UNIFORMITY SEMINORMS ON ℓ^∞ AND APPLICATIONS

BERNARD HOST AND BRYNA KRA

Abstract. A key tool in recent advances in understanding arithmetic progressions and other patterns in subsets of the integers is certain norms or seminorms. One example is the norms on $\mathbb{Z}/N\mathbb{Z}$ introduced by Gowers in his proof of Szemerédi's Theorem, used to detect uniformity of subsets of the integers. Another example is the seminorms on bounded functions in a measure preserving system (associated to the averages in Furstenberg's proof of Szemerédi's Theorem) defined by the authors. For each integer $k \geq 1$, we define seminorms on $\ell^\infty(\mathbb{Z})$ analogous to these norms and seminorms. We study the correlation of these norms with certain algebraically defined sequences, which arise from evaluating a continuous function on the homogeneous space of a nilpotent Lie group on a orbit (the nilsequences). Using these seminorms, we define a dual norm that acts as an upper bound for the correlation of a bounded sequence with a nilsequence. We also prove an inverse theorem for the seminorms, showing how a bounded sequence correlates with a nilsequence. As applications, we derive several ergodic theoretic results, including a nilsequence version of the Wiener-Wintner ergodic theorem, a nil version of a corollary to the spectral theorem, and a weighted multiple ergodic convergence theorem.

1. Introduction

1.1. **Norms and seminorms.** In his proof of Szemerédi's Theorem, Gowers [G] introduced norms for functions defined on $\mathbb{Z}/N\mathbb{Z}$ that count parallelepiped configurations and can be used to detect certain patterns (such as arithmetic progressions) in subsets of the integers. In [HK1], we defined seminorms on bounded measurable functions on a measure preserving system, that can be viewed as averages over parallelepipeds and use them to control the norm of multiple ergodic averages (such as one evaluated along arithmetic progressions). Although the original definitions were quite different, it turns out that the Gowers norms and the ergodic seminorms are almost the same object, but are defined on different spaces: one on the space of functions on $\mathbb{Z}/N\mathbb{Z}$ and the other on the space of bounded functions on a measure space. We used the ergodic seminorms to define factors of a measure space, and then showed that these factors have algebraic structure. This algebraic structure is the main ingredient in proving convergence of multiple ergodic averages along arithmetic progressions, and along other sequences. Gowers norms have since been used in other contexts, including the proof of Green and Tao [GT1] that the primes contain arbitrarily long arithmetic progressions. The connection between nilsystems in ergodic theory and the algebraic nature of analogous combinatorial objects has yet to be fully understood. The beginning of this is carried out by Green and Tao (see [GT2], [GT3] and [GT4]), including an inverse theorem for the third Gowers norm.

The second author was partially supported by NSF grant DMS-0555250.

In this article, we define related seminorms on bounded sequences and prove a structure theorem and an inverse theorem for it. We also give some ergodic theoretic applications of these constructions. These applications include a version of the Wiener-Wintner ergodic theorem extended to nilsequences, a spectral type theorem for nilsequences, and a weighted ergodic theorem. Polynomial versions of these results are contained in a forthcoming article. All these properties depend on the connection to algebraic structures and we describe these structures more precisely.

1.2. **Nilsystems and nilsequences.** In the inverse and structure theorems described above, a key role is played by algebraic objects, the nilsystems:

Definition 1.1. Assume that G is a k-step nilpotent Lie group and $\Gamma \subset G$ is a discrete, cocompact subgroup of G. The compact manifold $X = G/\Gamma$ is called a *k-step nilmanifold*. The *Haar measure* μ of X is the unique probability measure invariant under the action $x \mapsto g.x$ of G on X by left translations. Letting T denote left multiplication by the fixed element $\tau \in G$, we call (X, μ, T) a *k-step nilmanifold*[1].

Loosely speaking, the Structure Theorem of [HK1] states that if one wants to understand the multiple ergodic averages

$$\frac{1}{N}\sum_{n=0}^{N-1} f_1(T^n x)\dots f_k(T^{kn}x) \ ,$$

where $k \geq 1$ is an integer, (X, μ, T) is a measure preserving system, and $f_1, \dots, f_k \in L^\infty(\mu)$, one can replace each function by its conditional expectation on some nilsystem. Thus one can reduce the problem to studying the same average in a nilsystem, reducing averaging in an arbitrary system to a more tractable question.

A related problem is study of the multicorrelation sequence

$$c_n := \int T^n f \cdot T^{2n} f \cdot \dots \cdot T^{kn} f \, d\mu \ ,$$

where $k \geq 1$ is an integer, (X, μ, T) is a measure preserving system, and $f \in L^\infty(\mu)$. In [BHK], we defined sequences that arise from nilsystems (the *nilsequences*) and show that a multicorrelation sequence can be decomposed into a sequence that is small in terms of density and a k-step nilsequence. We define this second term precisely:

Definition 1.2. Let (X, μ, T) be a k-step nilsystem, $f\colon X \to \mathbb{C}$ a continuous function, $\tau \in G$, and $x_0 \in X$. The sequence $(f(\tau^n x_0)\colon n \in \mathbb{Z})$ is a *basic k-step nilsequence*. If, in addition, the function f is smooth, then the sequence $(f(\tau^n x_0)\colon n \in \mathbb{Z})$ is called a *smooth k-step nilsequence*. A *k-step nilsequence* is a uniform limit of basic k-step nilsequences.

The family of k-step nilsequences forms a closed, shift invariant subalgebra of sequences in $\ell^\infty(\mathbb{Z})$. One step nilsequences are exactly the almost periodic sequences. An example of a 2-step nilsequence is the sequence $(\exp(\pi i n(n-1)\alpha)\colon n \in \mathbb{Z})$, where α lies in the torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. (The collection of all 2-step nilsequences is described fully and classified in [HK2].)

[1] X is endowed with its Borel σ-algebra $\mathcal{X}$. In general, we omit the associated σ-algebra from our notation, writing (X, μ, T) for a measure preserving probability system rather than $(X, \mathcal{X}, \mu, T)$. We implicitly assume that all measure preserving systems are probability systems.

1.3. **Direct theorems and inverse theorems.** We define a new seminorm on bounded sequences and use this seminorm, an associated dual norm, and nilsequences to derive direct and inverse theorems. These seminorms on $\ell^\infty(\mathbb{Z})$ arise via an averaging process, and there is more than one natural way to take such an average. The first is looking along a particular sequence of intervals of integers whose lengths tend to infinity, and taking the average over these intervals. This corresponds, in some sense, to a local point of view, as such an averaging scheme does not take into account what happens outside this particular sequence of intervals. A second way to take an average is to allow all choices of intervals. This uniform point of view gives us further information on the original sequence.

Averaging in $\mathbb{Z}$, the first version gives rise to the classic notion of density, taking the proportion of a set relative to the sequence of intervals $[1, \dots, N]$, while the second gives rise to the slightly different notion of Banach density, where the density is computed relative to any sequence of intervals whose lengths tend to infinity. Each type of averaging gives rise for each integer $k \geq 1$ to some sort of uniformity measurement (seminorm, norm, or a version thereof) on bounded sequences.

We use the seminorms associate to each of these averaging methods to address analogs of combinatorial results. A classical problem in combinatorics is to start with a finite set A of integers (for example) and say something about properties of sets that can be built from A, such as the sumset $A+A$ or product set $A \cdot A$. Such results are referred to as direct theorems. Inverse theorems start with the sumset, product set, or other information derived from a finite set, and then try to deduce information about the set itself.

We prove both a direct theorem and an inverse theorem. For the direct theorem, we show that there is a dual norm that acts as an upper bound on the correlation of a bounded sequence with a nilsequence. We also prove an inverse theorem for the seminorms, showing how a bounded sequence correlates with a nilsequence. This is an ℓ^∞ version of the Gowers Inverse Conjecture made by Green and Tao [GT3]. This conjecture was resolved by them for the third Gowers norm in [GT4].

Using the direct theorems, we derive a weighted multiple ergodic convergence theorem. We believe that one should be able to use these methods to derive other combinatorial results.

The tools used in this paper have several sources. One is a version of the Furstenberg Correspondence Principle (see [F]), used to translate the problems into ergodic theoretic statements. Another is the connection of the seminorms we define with the algebraic structure of nilsystems, using properties of the ergodic seminorms developed in [HK1]. Throughout, we use some harmonic analysis on nilmanifolds.

This article can be viewed as an ergodic perspective on the development of a "higher order Fourier analysis" that has been proposed by Green and Tao [GT3]. Our direct results develop harmonic analysis relative to the standard Fourier analytic methods and our local inverse results lend support to Green-Tao conjecture of an inverse theorem for the Gowers norms.

1.4. **Organization of the paper.** In the next section, we define the seminorms on $\ell^\infty(\mathbb{Z})$ and give their basic properties. We then state the main results first for $k = 2$ and then for general k, with the intention of clarifying the objects under study. Section 3 gives the background on ergodic seminorms and nilsystems. In Section 4, we give a presentation of the Correspondence Principle that allows us to prove the properties of the $\ell^\infty(\mathbb{Z})$ seminorms introduced in Section 2. In Section 5, we study the dual norm associated to these seminorms and use it to prove the direct theorems on the seminorms. We prove the inverse theorems in Section 6, using an extension of the Correspondence Principle and in Section 7 we give some ergodic theoretic consequences of these results. Throughout we make use of the connection with the ergodic seminorms.

2. Summary of the results

We introduce seminorms on $\ell^\infty(\mathbb{Z})$ corresponding to the Gowers norms [G] in the finite setting and to the seminorms in ergodic theory introduced in [HK1]. We begin with some definitions and statements of the main properties. After defining the relevant seminorms, we give the statements of the results, beginning with the sample case of $k = 2$.

Notation. We write sequences as $\mathbf{a} = (a_n : n \in \mathbb{Z})$ and we write the uniform norm of this sequence as $\|\mathbf{a}\|_\infty$.

By an *interval*, we mean an interval in $\mathbb{Z}$. If I is an interval, $|I|$ denotes its length.

We write $z \mapsto Cz$ for complex conjugation in $\mathbb{C}$. Thus $C^k z = z$ if k is an even integer and $C^k z = \bar{z}$ if k is an odd integer.

For every $k \geq 1$, points of $\mathbb{Z}^k$ are written $h = (h_1, \ldots, h_k)$. For $\epsilon = (\epsilon_1, \ldots, \epsilon_k) \in \{0,1\}^k$ and $h = (h_1, \ldots, h_k) \in \mathbb{Z}^k$, we define

$$|\epsilon| = \epsilon_1 + \ldots + \epsilon_k \text{ and } \epsilon \cdot h = \epsilon_1 \cdot h_1 + \ldots + \epsilon_k \cdot h_k \ .$$

Further notation on averages of sequences of intervals is given at the end of this Section.

2.1. The local "seminorms" and the uniformity seminorms on $\ell^\infty(\mathbb{Z})$.

We define two quantities that are measurements on bounded sequences. The proofs rely on material from a variety of sources (summarized in Section 3) and some machinery that we develop, and so we postpone them until Section 4. In fact, some of the properties stated in this section can be proved via direct computations. However, we prefer proofs relying on the Furstenberg correspondence principle, as we use a modification of this principle to prove stronger results.

We introduce the property that allows us to define certain "seminorms."

Definition 2.1. Let $k \geq 1$ be an integer, $\mathbf{a} = (a_n : n \in \mathbb{Z})$ be a bounded sequence, and $\mathbf{I} = (I_j : j \geq 1)$ be a sequence of intervals whose lengths tend to infinity. We say that the sequence $\mathbf{a}$ *satisfies property* $\mathcal{P}(k)$ *on* $\mathbf{I}$ if for all $h = (h_1, \ldots, h_k) \in \mathbb{Z}^k$, the limit

$$\lim_{j \to +\infty} \frac{1}{|I_j|} \sum_{n \in I_j} \prod_{\epsilon \in \{0,1\}^k} C^{|\epsilon|} a_{n + h \cdot \epsilon}$$

exists. We denote this limit by $c_h(\mathbf{I}, \mathbf{a})$.

Given a bounded sequence $\mathbf{a}$ and a sequence of intervals whose lengths tend to infinity, one can always pass to a subsequence on which $\mathbf{a}$ satisfies $\mathcal{P}(k)$.

Proposition 2.2. *Let $k \geq 1$ be an integer, $\mathbf{I} = (I_j : j \geq 1)$ be a sequence of intervals whose lengths tend to infinity, and let $\mathbf{a}$ be a bounded sequence satisfying* *property $\mathcal{P}(k)$ on $\mathbf{I}$. Then then limit*

$$\lim_{H\to+\infty}\frac{1}{H^k}\sum_{h_1,\dots,h_k=0}^{H-1} c_h(\mathbf{I},\mathbf{a})\ ,$$

exists and is non-negative.

Using this proposition, we define:

Definition 2.3. For an integer $k\geq 1$, a sequence of intervals $\mathbf{I}=(I_j\colon j\geq 1)$, and a bounded sequence $\mathbf{a}$ satisfying property $\mathcal{P}(k)$ on $\mathbf{I}$, define

$$\|\mathbf{a}\|_{\mathbf{I},k}=\Bigl(\lim_{H\to+\infty}\frac{1}{H^k}\sum_{h_1,\dots,h_k=0}^{H-1} c_h(\mathbf{I},\mathbf{a})\Bigr)^{1/2^k}\ .$$

We call $\|\cdot\|_{\mathbf{I},k}$ a *local "seminorm"* (with quotes on the word seminorm), because the space of sequences satisfying property $\mathcal{P}(k)$ on $\mathbf{I}$ is not a vector space. On the other hand, we do have:

Proposition 2.4. *Assume that $k\geq 1$ is an integer, $\mathbf{a}$ and $\mathbf{b}$ are bounded sequences, and $\mathbf{I}$ is a sequence of intervals whose lengths tend to infinity. If $\mathbf{a},\mathbf{b}$ and $\mathbf{a}+\mathbf{b}$ satisfy property $\mathcal{P}(k)$ on $\mathbf{I}$, then $\|\mathbf{a}+\mathbf{b}\|_{\mathbf{I},k}\leq\|\mathbf{a}\|_{\mathbf{I},k}+\|\mathbf{b}\|_{\mathbf{I},k}$.*

The "seminorms" are also non-increasing with k:

Proposition 2.5. *If the bounded sequence $\mathbf{a}$ satisfies properties $\mathcal{P}(k)$ and $\mathcal{P}(k+1)$ on the sequence of intervals $\mathbf{I}$, then $\|\mathbf{a}\|_{\mathbf{I},k}\leq\|\mathbf{a}\|_{\mathbf{I},k+1}$.*

We use the "seminorm" to define a measure of uniformity (a uniformity seminorm) on bounded sequences:

Definition 2.6. Let $\mathbf{a}=(a_n\colon n\in\mathbb{Z})$ be a bounded sequence and let $k\geq 1$ be an integer. We define the *k-uniformity seminorm* $\|\mathbf{a}\|_{U(k)}$ to be the supremum of $\|\mathbf{a}\|_{\mathbf{I},k}$, where the supremum is taken over all sequences of intervals $\mathbf{I}$ on which $\mathbf{a}$ satisfies property $\mathcal{P}(k)$.

Using Proposition 2.4, by passing, if necessary, to subsequences of the sequences of intervals, we immediately deduce:

Proposition 2.7. *For every integer $k\geq 2$, $\|\cdot\|_{U(k)}$ is a seminorm on $\ell^\infty(\mathbb{Z})$.*

2.2. **Comments on the definitions.**

2.2.1. The definitions of $\|\mathbf{a}\|_{\mathbf{I},k}$ and $\|\mathbf{a}\|_{U(k)}$ are very similar to those of the Gowers norms introduced in [G] in the finite setting (meaning, for sequences indexed by $\mathbb{Z}/N\mathbb{Z}$). In the sequel, we establish analogs of properties of Gowers norms for the $\ell^\infty(\mathbb{Z})$ seminorms. The $\ell^\infty(\mathbb{Z})$ seminorms are also close relatives of the ergodic seminorms of [HK1]. In the sequel we show that this resemblance is not merely formal; the link between the $\ell^\infty(\mathbb{Z})$ seminorms and the ergodic seminorms is a basic tool of this paper.

2.2.2. It can be shown that in Proposition 2.2 the averages on $[0,H-1]^k$ can be replaced by averages on any sequence of "rectangles" $(I_{H,1}\times\dots I_{H,k}\colon H\geq 1)$, where $I_{H,j}$ is an interval for every $j\in\{1,\dots,k\}$ and every H and $\min_j|I_{H,j}|\to+\infty$ as $H\to+\infty$; more generally we could also average over any Følner sequence in $\mathbb{Z}^k$.

2.2.3. For clarity, we explain what the definitions mean when $k = 1$. (We discuss $k = 2$ in the next section.) Let $\mathbf{a} = (a_n\colon n \in \mathbb{Z})$ be a bounded sequence and let $\mathbf{I} = (I_j\colon j \geq 1)$ be a sequence of intervals whose lengths tend to infinity.

Property $\mathcal{P}(1)$ says that for every $h \in \mathbb{Z}$, the averages

$$\frac{1}{|I_j|}\sum_{n\in I_j} a_n\overline{a_{n+h}}$$

converge as $j \to +\infty$ and the definition of $\|\mathbf{a}\|_{\mathbf{I},1}$ is

$$\|\mathbf{a}\|_{\mathbf{I},1} = \Bigl(\lim_{H\to+\infty}\frac{1}{H}\sum_{h=0}^{H-1}\lim_{j\to+\infty}\frac{1}{|I_j|}\sum_{n\in I_j} a_n\overline{a_{n+h}}\Bigr)^{1/2} .$$

Furthermore,

$$\|\mathbf{a}\|_{\mathbf{I},1} \geq \limsup_{j\to+\infty}\Bigl|\frac{1}{|I_j|}\sum_{n\in I_j} a_n\Bigr|$$

and

$$\|\mathbf{a}\|_{U(1)} = \lim_{N\to+\infty}\ \sup_{M\in\mathbb{Z}}\Bigl|\frac{1}{N}\sum_{n=M}^{M+N-1} a_n\Bigr| .$$

The first property follows easily from the van der Corput Lemma (see Appendix A) and probably the second can also be proved directly. Both properties also follow from the discussion in Section 4.2.

2.2.4. The difference between the local "seminorms" and the uniformity seminorms is best illustrated by considering a randomly generated sequence. Let $\mathbf{a} = (a_n\colon n \in \mathbb{Z})$ be a random sequence, where the a_n are independent random variables, taking the values $+1$ and -1 each with probability $1/2$. Let $\mathbf{I} = (I_j\colon j \geq 1)$ be a sequence of intervals whose lengths tend to infinity. Then for every integer k, the sequence $\mathbf{a}$ satisfies property $\mathcal{P}(k)$ on $\mathbf{I}$ almost surely and $\|\mathbf{a}\|_{\mathbf{I},k} = 0$. On the other hand, we have that $\|\mathbf{a}\|_{U(k)} = 1$ almost surely. Indeed, for every integer $j \geq 1$ there exists an interval I_j of length j on which the sequence $\mathbf{a}$ is constant and equal to 1; taking $\mathbf{I}$ to be this sequence of intervals, we have that $\|\mathbf{a}\|_{\mathbf{I},k} = 1$ for every integer $k \geq 1$. The apparent contradiction only arises because of the choice of uncountably many sequences of intervals.

2.2.5. There are nontrivial bounded sequences for which the uniformity seminorm is 0. This is illustrated by the following particular case of Corollary 3.11.

Proposition 2.8. *Let $k \geq 1$ be an integer and assume that (X, T) is a uniquely ergodic system with invariant measure μ) that is weakly mixing. If f is a function on X with $\int f\,d\mu = 0$, then for every $x \in X$, the sequence $(f(T^n x)\colon n \in \mathbb{Z})$ has 0 k-uniformity seminorm.*

2.3. **The case $k = 2$.** To further clarify the statements, we explain some of our general results in the particular case that $k = 2$. These results are prototypes for the general case, but are simpler to state and prove. Most of these results can be proved without resorting to any significant machinery and we include one of the simpler proofs here.

Notation. We write $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. For $t \in \mathbb{T}$, $e(t) = \exp(2\pi i t)$.

The first result explains the role of the local "seminorm", namely that it acts as an upper bound:

Proposition 2.9. *If* $\mathbf{a} = (a_n \colon n \in \mathbb{Z})$ *is a bounded sequence satisfying* $\mathcal{P}(2)$ *on the sequence of intervals* $\mathbf{I} = (I_j \colon j \geq 1)$*, then*

$$\limsup_{j\to+\infty} \sup_{t\in\mathbb{T}} \Big| \frac{1}{|I_j|} \sum_{n\in I_j} a_n e(nt) \Big| \leq \|\mathbf{a}\|_{\mathbf{I},2} \ .$$

Proof. We can assume that $\|\mathbf{a}\|_\infty \leq 1$. By the van der Corput Lemma (Appendix A), Cauchy-Schwartz Inequality, and another application of the van der Corput Lemma, we have that for all integers $j, H \geq 1$, and all $t \in \mathbb{T}$,

$$\begin{aligned} \Big| \frac{1}{|I_j|} \sum_{n\in I_j} a_n e(nt) \Big|^4 &\leq \Big(\frac{cH}{|I_j|} + \Big| \sum_{h=-H}^{H} \frac{H-|h|}{H^2} \frac{1}{|I_j|} \sum_{n\in I_j} a_n \overline{a_{n+h}} \Big| \Big)^2 \\ &\leq \frac{c'H}{|I_j|} + \sum_{h=-H}^{H} \frac{H-|h|}{H^2} \Big| \frac{1}{|I_j|} \sum_{n\in I_j} a_n \overline{a_{n+h}} \Big|^2 \\ &\leq \frac{c''H}{|I_j|} + \sum_{\ell=-H}^{H} \sum_{h=-H}^{H} \frac{H-|\ell|}{H^2} \frac{H-|h|}{H^2} \Big| \frac{1}{|I_j|} \sum_{n\in I_j} a_n \overline{a_{n+h}}\, \overline{a_{n+\ell}} a_{n+h+\ell} \Big| \ , \end{aligned}$$

where c, c', c'' are universal constants. Taking the limit as $j \to +\infty$ first (recall that the sequence $\mathbf{a}$ satisfies $\mathcal{P}(2)$ on the sequence of intervals $\mathbf{I}$), and then as $H \to +\infty$, we have the announced result. □

We use this to show how such a sequence $\mathbf{a}$ correlates with almost periodic sequences. First a definition:

Definition 2.10. A sequence of the form $(e(nt) \colon n \in \mathbb{Z})$ is called a *complex exponential sequence*. A sequence is a *trigonometric polynomial* if it is a finite linear combination of complex exponential sequences. An *almost periodic sequence* is a uniform limit of trigonometric polynomials.

By approximation, it follows immediately from Proposition 2.9 that:

Corollary 2.11. *Let* $\mathbf{b} = (b_n \colon n \in \mathbb{Z})$ *be an almost periodic sequence. Then for every* $\delta > 0$*, there exists a constant* $c = c(\mathbf{b}, \delta)$ *such that if a bounded sequence* $\mathbf{a} = (a_n \colon n \in \mathbb{Z})$ *satisfies property* $\mathcal{P}(2)$ *on a sequence of intervals* $\mathbf{I} = (I_j \colon j \geq 1)$*, then*

$$\limsup_{j\to+\infty} \Big| \frac{1}{|I_j|} \sum_{n\in I_j} a_n b_n \Big| \leq c\|\mathbf{a}\|_{\mathbf{I},2} + \delta\|\mathbf{a}\|_\infty \ .$$

For some almost periodic sequences we have more precise bounds. A smooth almost periodic sequence $\mathbf{b} = (b_n \colon n \in \mathbb{Z})$ (that is, a smooth 1-step nilsequence) can be written as

$$b_n = \sum_{m=1}^{\infty} \lambda_m e(nt_m) \ ,$$

where t_m, $m \geq 1$, are distinct elements of $\mathbb{T}$ and $\lambda_m \in \mathbb{C}$, $m \geq 1$, satisfy

$$\sum_{m=1}^{\infty} |\lambda_m| < +\infty \ .$$

We define

$$\|\!|\mathbf{b}|\!\|_2^* = \Big(\sum_{m=1}^{\infty} |\lambda_m|^{4/3}\Big)^{3/4}$$

and we have that:

Proposition 2.12. *Let $\mathbf{a} = (a_n : n \in \mathbb{Z})$ be a bounded sequence satisfying property $\mathcal{P}(2)$ on the sequence of intervals $\mathbf{I} = (I_j : j \geq 1)$ and $\mathbf{b} = (b_n : n \in \mathbb{Z})$ be a smooth almost periodic sequence. Then,*

$$\limsup_{j\to+\infty}\Big|\frac{1}{|I_j|}\sum_{n\in I_j} a_n b_n\Big| \leq \|\mathbf{a}\|_{\mathbf{I},2}\, \|\!|\mathbf{b}|\!\|_2^* \ .$$

The constant $\|\!|\mathbf{b}|\!\|_2^*$ here is the best possible. Undoubtedly, one could prove this result without resorting to special machinery, but we do not attempt this method as this is a particular case of a general result (Theorem 2.13). In fact we show that the norm $\|\!|\cdot|\!\|_2^*$ acts as the dual of the seminorm $\|\cdot\|_{U(2)}$.

2.4. **Main results.** Let $k \geq 2$ be an integer. In section 5.3, for every $(k-1)$-step nilmanifold X we define a norm $\|\!|\cdot|\!\|_k^*$ on the space $\mathcal{C}^\infty(X)$ of smooth functions on X. We defer the precise definition, as it requires development of some further background. Let $\mathbf{b}$ be a smooth $(k-1)$-step nilsequence. Then there exists an ergodic $(k-1)$-step nilsystem (Corollary 3.3), a smooth function f on x, and a point $x_0 \in X$ with

$$b_n = f(T^n x_0) \text{ for every } n \in \mathbb{Z} \ .$$

The same sequence $\mathbf{b}$ can be represented in this way in several manners, with different systems, different starting points, and different functions, but we show (Corollary 5.8) that all associated functions f have the same norm $\|\!|\cdot|\!\|_k^*$. Therefore we can define $\|\!|\mathbf{b}|\!\|_k^* = \|\!|f|\!\|_k^*$ where f is any of the possible functions.

2.4.1. *Direct results.* Using this norm, we have generalizations of the results already given for $k = 2$:

Theorem 2.13 (Direct Theorem)**.** *Let $\mathbf{a} = (a_n : n \in \mathbb{Z})$ be a bounded sequence that satisfies property $\mathcal{P}(k)$ on the sequence of intervals $\mathbf{I} = (I_j : j \geq 1)$. For all $(k-1)$-step smooth nilsequences $\mathbf{b}$, we have*

$$\limsup_{j\to+\infty}\Big|\frac{1}{|I_j|}\sum_{n\in I_j} a_n b_n\Big| \leq \|\mathbf{a}\|_{\mathbf{I},k}\, \|\!|\mathbf{b}|\!\|_k^* \ .$$

By density, Theorem 2.13 immediately implies:

Corollary 2.14. *Let $\mathbf{b} = (b_n : n \in \mathbb{Z})$ be a $(k-1)$-step nilsequence and $\delta > 0$. There exists a constant $c = c(\mathbf{b}, \delta)$ such that for every bounded sequence $\mathbf{a} = (a_n : n \in \mathbb{Z})$ satisfying property $\mathcal{P}(k)$ on a sequence of intervals $\mathbf{I} = (I_j : j \geq 1)$, we have*

$$\limsup_{j\to+\infty}\Big|\frac{1}{|I_j|}\sum_{n\in I_j} a_n b_n\Big| \leq c\|\mathbf{a}\|_{\mathbf{I},k} + \delta\|\mathbf{a}\|_\infty \ .$$

Using these results, we immediately deduce uniform versions:

Corollary 2.15. *Let $\mathbf{b} = (b_n : n \in \mathbb{Z})$ be a smooth $(k-1)$-step nilsequence and $\mathbf{a} = (a_n : n \in \mathbb{Z})$ be a bounded sequence. Then*

$$\lim_{N\to+\infty} \sup_{M\in\mathbb{Z}} \Big| \frac{1}{N} \sum_{n=M}^{N+M-1} a_n b_n \Big| \leq \|\mathbf{a}\|_{U(k)} \, \|\!|\mathbf{b}|\!\|_k^* \ .$$

Let $\mathbf{b} = (b_n : n \in \mathbb{Z})$ be a $(k-1)$-step nilsequence and let $\delta > 0$. There exists a constant $c = c(\mathbf{b}, \delta)$ such that for every bounded sequence $\mathbf{a} = (a_n : n \in \mathbb{Z})$,

$$\lim_{N\to+\infty} \sup_{M\in\mathbb{Z}} \Big| \frac{1}{N} \sum_{n=M}^{N+M-1} a_n b_n \Big| \leq c\|\mathbf{a}\|_{U(k)} + \delta\|\mathbf{a}\|_\infty \ .$$

We refer to these results as direct results, meaning that we start with a sequence and derive its correlation with nilsequences. One can view them as upper bounds, because they give an upper bound between the correlation of a sequence with a nilsequence.

2.4.2. *Inverse results.* The next results are in the opposite direction of the direct results of the previous section, and we refer to them as “inverse results”.

Theorem 2.16 (Inverse Theorem)**.** *Let $\mathbf{a} = (a_n : n \in \mathbb{Z})$ be a bounded sequence. Then for every $\delta > 0$, there exists a $(k-1)$-step smooth nilsequence $\mathbf{b} = (b_n : n \in \mathbb{Z})$ such that*

$$\|\!|\mathbf{b}|\!\|_k^* = 1 \text{ and } \lim_{N\to+\infty} \sup_{M\in\mathbb{Z}} \Big| \frac{1}{N} \sum_{n=M}^{M+N-1} a_n b_n \Big| \geq \|\mathbf{a}\|_{U(k)} - \delta \ .$$

Summarizing this theorem and Corollary 2.15 we have

Corollary 2.17. *For every bounded sequence $\mathbf{a} = (a_n : n \in \mathbb{Z})$,*

$$\|\mathbf{a}\|_{U(k)} = \sup_{\substack{\mathbf{b}=(b_n) \text{ is a smooth} \\ \text{nilsequence and } \|\!|\mathbf{b}|\!\|_k^*=1}} \lim_{N\to+\infty} \sup_{M\in\mathbb{Z}} \Big| \frac{1}{N} \sum_{n=M}^{N+M-1} a_n b_n \Big| \ .$$

This means that we can view the norm $\|\!|\cdot|\!\|_k^*$ as the dual norm of the uniformity seminorm $\|\cdot\|_{U(k)}$.

Corollary 2.18. *For a bounded sequence $\mathbf{a} = (a_n : n \in \mathbb{Z})$, the following properties are equivalent:*

(i) $\|\mathbf{a}\|_{U(k)} = 0$.

(ii) $\displaystyle\lim_{N\to+\infty} \sup_{M\in\mathbb{Z}} \Big| \frac{1}{N} \sum_{n=M}^{N+M-1} a_n b_n \Big| = 0$ *for every $(k-1)$-step smooth nilsequence $\mathbf{b} = (b_n : n \in \mathbb{Z})$.*

(iii) $\displaystyle\lim_{N\to+\infty} \sup_{M\in\mathbb{Z}} \Big| \frac{1}{N} \sum_{n=M}^{N+M-1} a_n b_n \Big| = 0$ *for every $(k-1)$-step nilsequence $\mathbf{b} = (b_n : n \in \mathbb{Z})$.*

For $k = 2$, Corollary 2.18, Proposition 2.9, and a density argument imply that the three equivalent conditions of Corollary 2.18 are also equivalent to

(iv) For every $t \in \mathbb{T}$, $\displaystyle\lim_{N\to+\infty} \sup_{M\in\mathbb{Z}} \Big| \frac{1}{N} \sum_{n=M}^{N+M-1} a_n e(nt) \Big| = 0.$

(v) $\displaystyle\lim_{N\to+\infty}\sup_{t\in\mathbb{T}}\sup_{M\in\mathbb{Z}}\Bigl|\frac{1}{N}\sum_{n=M}^{M+N-1}a_n e(nt)\Bigr|=0.$

2.4.3. *A counterexample.* It is important to note that the inverse results have no version involving local "seminorms" and we give here an example illustrating this point.

Let $(N_j\colon j\geq 1)$ be an increasing sequence of integers with $N_1=0$ and tending sufficiently fast to $+\infty$. For $j\geq 1$ let $I_j=[N_j,N_{j+1}-1]$ and let $\mathbf{I}=(I_j\colon j\geq 1)$. Let the sequence $\mathbf{a}$ be defined by $a_n=e(n/j)$ if $N_j\leq |n|<N_{j+1}$. Then $\|\mathbf{a}\|_{\mathbf{I},2}=1$ and for every $t\in\mathbb{T}$, the average of $a_ne(nt)$ on the interval I_j converges to zero as $j\to+\infty$. Therefore, for every almost periodic sequence $\mathbf{b}$, the average of a_nb_n on I_j also converges to zero.

This highlights a difference between the finite case, where the norms are defined on $\mathbb{Z}/N\mathbb{Z}$, and the infinite case. One can not construct such a sequence where the behavior worsens as one tends to infinity.

2.5. **A condition for convergence.**

Theorem 2.19. *For a bounded sequence $\mathbf{a}=(a_n\colon n\in\mathbb{Z})$, the following are equivalent.*

(i) *For every $\delta>0$, the sequence $\mathbf{a}$ can be written as $\mathbf{a}'+\mathbf{a}''$ where $\mathbf{a}'$ is a $(k-1)$-step nilsequence and $\|\mathbf{a}''\|_{U(k)}<\delta$.*

(ii) *For every $(k-1)$-step nilsequence $\mathbf{c}=(c_n\colon n\in\mathbb{Z})$, the averages of a_nc_n converge, meaning that the limit*
$$\lim_{j\to+\infty}\frac{1}{|I_j|}\sum_{n\in I_j}a_nc_n$$
exists for every sequence $(I_j\colon j\geq 1)$ of intervals whose lengths tend to infinity.

In Proposition 7.1, we give a method to build sequences satisfying the (equivalent) properties of Theorem 2.19, checking that the sequences verify the first property. As this proposition uses material not yet defined, we do not state it here but only give two examples of its application.

A generalized polynomial is defined to be a real valued function that is obtained from the identity function and real constants by using (in arbitrary order) the operations of addition, multiplication, and taking the integer part. We have:

Proposition 2.20. *Let p be a generalized polynomial and for every $n\in\mathbb{Z}$, let $\{p(n)\}$ be the fractional part of $p(n)$. Then the sequences $(\{p(n)\}\colon n\in\mathbb{Z})$ and $(e(p(n))\colon n\in\mathbb{Z})$ satisfy the (equivalent) properties of Theorem 2.19.*

The *Thue-Morse sequence* $\mathbf{a}=(a_n\colon n\in\mathbb{Z})$ is given by $a_n=1$ if the sum of the digits of $|n|$ written in base 2 is odd and $a_n=0$ otherwise. In Section 7.2 we show:

Proposition 2.21. *The Thue-Morse sequence satisfies the properties of Theorem 2.19.*

A similar method can be used for other sequences, for example for all sequences associated to primitive substitutions of constant length (see [Q] for the definition).

2.6. **An application to ergodic theory.**

2.6.1. We recall a classical result in ergodic theory.

Theorem (Wiener-Wintner ergodic theorem [WW]). *Let (X,μ,T) be an ergodic system and $\phi\in L^\infty(\mu)$. Then there exists $X_0\subset X$ with $\mu(X_0)=1$ such that*

$$\frac{1}{N}\sum_{n=0}^{N-1}\phi(T^nx)\,e(nt)$$

converges for every $x\in X_0$ and every $t\in\mathbb{T}$.

The important point here is that the set X_0 does not depend on the choice of t.

We also recall an immediate corollary of the spectral theorem:

Corollary (of the spectral theorem). *Let $\mathbf{a}=(a_n\colon n\in\mathbb{Z})$ be a bounded sequence and assume that*

$$\lim_{N\to+\infty}\sum_{n=0}^{N-1}a_ne(nt)$$

exists for every $t\in\mathbb{T}$. Then for every system (Y,ν,S) and every $f\in L^2(\nu)$, the averages

$$\frac{1}{N}\sum_{n=0}^{N-1}a_nS^nf$$

converge in $L^2(\nu)$ as $N\to+\infty$.

Putting these two results together, we have:

Corollary. *Assume that (X,μ,T) is an ergodic system and $\phi\in L^\infty(\mu)$. There exists $X_0\subset X$ with $\mu(X_0)=1$ such that for every $x\in X_0$, every system (Y,ν,S), and every $f\in L^2(\nu)$, the averages*

$$\frac{1}{N}\sum_{n=0}^{N-1}\phi(T^nx)S^nf$$

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

The strength of this result is that the set X_0 does not depend either on Y or on f. We say that for every $x\in X_0$, the sequence $(\phi(T^nx))$ is a *universally good for the convergence in mean of ergodic averages.* In fact, for almost every x, this sequence is also universally good for the almost everywhere convergence [BFKO], but we do not address this strengthening here.

2.6.2. We generalize these results for multiple ergodic averages. We start with a generalization of the Wiener-Wintner Theorem, where we can replace the exponential sequence $e(nt)$ by an arbitrary nilsequence.

Theorem 2.22 (A generalized Wiener-Wintner Theorem). *Let (X,μ,T) be an ergodic system and ϕ be a bounded measurable function on X. Then there exists $X_0\subset X$ with $\mu(X_0)=1$ such that for every $x\in X_0$, the averages*

$$\frac{1}{N}\sum_{n=0}^{N-1}\phi(T^nx)\,b_n$$

converge as $N\to+\infty$ for every $x\in X_0$ and every nilsequence $\mathbf{b}=(b_n\colon n\in\mathbb{Z})$.

We give a sample application. Generalized polynomials were defined in Section 2.5.

Corollary 2.23. *Let (X,μ,T) be an ergodic system, ϕ be a bounded measurable function on X, and X_0 be the subset of X introduced in Theorem 2.22. Then for every $x\in X_0$ and every generalized polynomial p, the averages*

$$\frac{1}{N}\sum_{n=0}^{N-1}\phi(T^nx)\{p(n)\}\ \text{and}\ \frac{1}{N}\sum_{n=0}^{N-1}\phi(T^nx)e(p(n))$$

converge.

(Recall that $\{p(n)\}$ denotes the fractional part of $p(n)$.) For standard polynomial sequences, this result was proven by Lesigne [Les2].

We next have a version of the spectral result for higher order nilsequences:

Theorem 2.24 (A substitute for the corollary of the Spectral Theorem)**.** *Let $k\geq 1$ be an integer and $\mathbf{a}=(a_n\colon n\in\mathbb{Z})$ be a bounded sequence such that the averages*

$$\frac{1}{N}\sum_{n=0}^{N-1}a_nb_n$$

converge as $N\to+\infty$ for every k-step nilsequence $\mathbf{b}=(b_n\colon n\in\mathbb{Z})$. Then for every system (Y,ν,S) and every $f_1,\dots,f_k\in L^\infty(\nu)$, the averages

$$\frac{1}{N}\sum_{n=0}^{N-1}a_n\,S^nf_1.S^{2n}f_2.\cdots.S^{kn}f_k \tag{1}$$

converge in $L^2(\nu)$.

Combining these theorems, we immediately deduce:

Theorem 2.25. *Let (X,μ,T) be an ergodic system and $\phi\in L^\infty(\mu)$. Then there exists $X_0\subset X$ with $\mu(X_0)=1$ such that for every $x_0\in X$, every system (Y,ν,S), every integer $k\geq 1$, and all functions $f_1,\dots,f_k\in L^\infty(\nu)$, the averages*

$$\frac{1}{N}\sum_{n=0}^{N-1}\phi(T^nx)\,S^nf_1.S^{2n}f_2.\cdots.S^{kn}f_k$$

converge in $L^2(\nu)$ as $N\to+\infty$.

In short, for every $x\in X_0$, the sequence $(\phi(T^nx))$ is *universally good for the convergence in mean of multiple ergodic averages.*

While Theorems 2.22 and 2.24 are results about nilsequences, nilsequences do not appear in the statement of Theorem 2.25: they occur only as tools in the proof, playing the role of complex exponentials in the classical results.

By successively using Theorems 2.19 and 2.24, we obtain further examples of universally good sequences for the convergence in mean of multiple ergodic averages. For example, by Proposition 2.20, for every generalized polynomial p the sequence $(\{p(n)\}\colon n\in\mathbb{Z})$ is a universally good sequence for the convergence in mean of multiple ergodic averages, as is the sequence $(e(p(n))\colon n\in\mathbb{Z})$. By Proposition 2.21, so is the Thue-Morse sequence.

2.7. **Some notation for averages.** In this paper we continuously take limits of averages on sequences of intervals. Writing the cumbersome formulas or replacing them by long explanations would make the paper unreadable and so we introduce some short notation. However, we continue using explicit formulas in the main statements.

We have several different notions of averaging for a sequence in $\ell^\infty(\mathbb{Z})$: over a particular sequence of intervals or uniformly over all intervals.

For averaging over a particular sequence of intervals, we define:

Definition 2.26. Let $\mathbf{a} = (a_n : n \in \mathbb{Z})$ be a bounded sequence and let $\mathbf{I} = (I_j : j \geq 1)$ be a sequence of intervals whose lengths $|I_j|$ tend to infinity. Define

$$\operatorname{limsup} |\operatorname{averages}_{\mathbf{I}}(a_n)| = \limsup_{j\to+\infty} \Bigl| \frac{1}{|I_j|} \sum_{n\in I_j} a_n \Bigr| \ .$$

The *averages of the sequence* $\mathbf{a}$ *on* $\mathbf{I}$ *converge* if the limit

$$\lim_{j\to+\infty} \frac{1}{|I_j|} \sum_{n\in I_j} a_n$$

exists. We denote this limit by $\lim \operatorname{averages}_{\mathbf{I}}\bigl(a_n\bigr)$ and call this the *average over* $\mathbf{I}$ *of the sequence* $\mathbf{a}$.

For taking a uniform average, we define:

Definition 2.27. Let $\mathbf{a} = (a_n : n \in \mathbb{Z})$ be a bounded sequence. The *upper limit of the averages of the sequence* $\mathbf{a}$ is defined to be

$$\operatorname{limsup} |\operatorname{averages}(a_n)| = \lim_{N\to+\infty} \sup_{M\in\mathbb{Z}} \Bigl| \frac{1}{N} \sum_{n=M}^{M+N-1} a_n \Bigr| \ .$$

(Note that this limit exists by subadditivity.)

The *averages of the sequence* $\mathbf{a}$ *converge* if the limit $\lim \operatorname{averages}_{\mathbf{I}}\bigl(a_n\bigr)$ exists for all sequences of intervals $\mathbf{I} = (I_j : j \geq 1)$ whose lengths $|I_j|$ tend to infinity. We denote this (common) limit by $\lim \operatorname{averages}\ \bigl(a_n\bigr)$ and call this the *uniform average of the sequence* $\mathbf{a}$.

Assuming the existence of the uniform average, it follows that

$$\lim_{N\to+\infty} \sup_{M\in\mathbb{Z}} \Bigl| \lim \operatorname{averages}\ \bigl(a_n\bigr) - \frac{1}{N} \sum_{n=M}^{M+N-1} a_n \Bigr| = 0 \ .$$

3. Some tools

3.1. **Nilmanifolds and nilsystems.**

3.1.1. *The definitions.* Short definitions were given in the introduction and we repeat them here in a more complete form.

Let G be a group. For $g, h \in G$, we write $[g,h] = ghg^{-1}h^{-1}$ for the commutator of g and h and we write $[A,B]$ for the subgroup spanned by $\{[a,b] : a \in A, b \in B\}$. The *commutator subgroups* G_j, $j \geq 1$, are defined inductively by setting $G_1 = G$ and $G_{j+1} = [G_j, G]$. Let $k \geq 1$ be an integer. We say that G is *k-step nilpotent* if G_{k+1} is the trivial subgroup.

Let G be a k-step nilpotent Lie group and Γ a discrete cocompact subgroup of G. The compact manifold $X = G/\Gamma$ is called a *k-step nilmanifold*. The group G

acts on X by left translations and we write this action as $(g,x)\mapsto g.x$. The *Haar measure* μ of X is the unique probability measure on X invariant under this action.

Let $\tau\in G$ and T be the transformation $x\mapsto \tau.x$ of X. Then (X,T,μ) is called a *k-step nilsystem*. When the measure is not needed for results, we omit and write that (X,T) is a k-step nilsystem.

Nilsystems are *distal* topological dynamical systems. This means that, if d_X is a distance on X defining its topology, then for every $x,x'\in X$,

$$\text{if } x\neq x',\ \text{ then } \inf_{n\in\mathbb{Z}} d_X(T^ny,T^ny')>0\ .$$

Let f be a continuous (respectively, smooth) function on X and $x_0\in X$. The sequence $(f(T^nx_0)\colon n\in\mathbb{Z})$ is called a *basic* (respectively, *smooth*) *k-step nilsequence*. A k-step nilsequence is a uniform limit of basic k-step nilsequences. Therefore, smooth k-step nilsequences are dense in the space of all k-step nilsequences under the uniform norm.

The Cartesian product of two k-step nilsystems is again a k-step nilsystem. It follows that the space of k-step nilsequences is an algebra under pointwise addition and multiplication. Moreover, this algebra is invariant under the shift.

As an example, 1-step nilsystems are translations on compact abelian Lie groups and 1-step nilsequences are exactly almost periodic sequences. For examples of 2-step nilsystems and a detailed study of 2-step nilsequences, see [HK2].

A general reference on nilsystems is [AGH] and the results summarized in the next few sections are contained in the literature. See, for example [Les1] and [Lei].

3.1.2. *Ergodicity.*

Theorem 3.1. *Let $k\geq 1$ be an integer. For a k-step nilsystem $(X=G/\Gamma,T)$ with Haar measure μ, the following properties are equivalent:*

(i) *(X,T) is transitive, meaning that it admits a dense orbit.*
(ii) *(X,T) is minimal, meaning that every orbit is dense.*
(iii) *(X,T) is uniquely ergodic.*
(iv) *(X,μ,T) is ergodic.*

When these properties are satisfied, we say that the system is *ergodic*, even in statements of topological nature (that is, without mention of the measure).

Theorem 3.2. *Let $k\geq 1$ be an integer, $(X=G/\Gamma,T)$ be a k-step nilsystem where T is the translation by $\tau\in G$. Let $x_0\in X$ and let Y be the closed orbit of x_0, meaning that Y is the closure of the orbit $\{T^nx_0\colon n\in\mathbb{Z}\}$. Then (Y,T) is a k-step nilsystem. More precisely, there exist a closed subgroup G' of G containing τ, such that $\Gamma'=\Gamma\cap G'$ is cocompact in G' and $Y=G'/\Gamma'$.*

If $(f(T^nx_0)\colon n\in\mathbb{Z})$ is a basic (respectively, smooth) nilsequence, by substituting the closed orbit of x_0 for X, we deduce:

Corollary 3.3. *For every basic (respectively, smooth) k-step nilsequence $\mathbf{a}=(a_n\colon n\in\mathbb{Z})$, there exists an ergodic k-step nilsystem (X,T), $x_0\in X$, and a continuous (respectively, smooth) function f on X with $a_n=f(T^nx_0)$ for every $n\in\mathbb{Z}$.*

Corollary 3.4. *Let $\mathbf{a}=(a_n\colon n\in\mathbb{Z})$ be a nilsequence. Then the averages of $\mathbf{a}$ converge.*

Proof. By density, we can restrict to the case that $\mathbf{a}$ is a basic nilsequence, and we write it as in Corollary 3.3. By unique ergodicity of (X, T), the averages converge to $\int f\,d\mu$, where μ is the Haar measure of X. $\square$

3.1.3. *A criteria for ergodicity.*

Theorem 3.5. *Let $k \geq 1$ be an integer, $(X = G/\Gamma, T)$ be a k-step nilsystem, and assume that T is translation by $\tau \in G$. Assume that*

(*) The group G is spanned by the connected component G_0 of its unit and by τ.

Then (X, T) is ergodic if and only if the translation induced by τ on the compact abelian group $Z = G/G_2\Gamma$ is ergodic.

Conversely, let $(X = G/\Gamma, T)$ be an ergodic nilsystem where T is the translation by $\tau \in G$. Let G_1 be the subgroup spanned by G_0 and τ and set $\Gamma_1 = \Gamma \cap G_1$. Then G_1 is an open subgroup of G, Γ_1 is a discrete cocompact subgroup of G_1, and by ergodicity, the image of G_1 in X under the natural projection is onto. We can therefore identify X with G_1/Γ_1. Thus we can assume that hypothesis (*) of Theorem 3.5 is satisfied. Throughout this paper, we implicitly assume that this hypothesis holds.

3.1.4. *The case of several commuting transformations.* Let $X = G/\Gamma$ be a nilmanifold and let $\tau_1, \dots, \tau_\ell$ be commuting elements of G. For $1 \leq i \leq \ell$ let $T_i \colon X \to X$ be the translation by τ_i. Then the results of Section 3.1.2 still hold, modulo the obvious changes. We do not give the modified statements here, with the exception of Theorem 3.5:

Theorem 3.6. *Let $X = G/\Gamma$ be a nilmanifold, $\tau_1, \dots, \tau_\ell$ be commuting elements of G, and for $1 \leq i \leq \ell$ let $T_i \colon X \to X$ be the translation by τ_i. Assume that:*

(**) The group G is spanned by the connected component G_0 of its unit and by $\tau_1, \dots, \tau_\ell$.

Then X is ergodic under the action of $T_1, T_2, \dots, T_\ell$ if and only if the action induced by these transformations on the compact abelian group $Z = G/G_2\Gamma$ is ergodic.

3.2. **The measures $\mu^{[k]}$ and HK-seminorms.**

In the rest of this section we consider arbitrary ergodic systems and we assume that $k \geq 1$ is an integer. We review the construction and properties of certain objects on X^{2^k} defined in [HK1].

3.2.1. *Some notation.* We introduce some notation to keep track of the 2^k copies of X. If X is a set, we write $X^{[k]} = X^{2^k}$ and index these copies of X by $\{0,1\}^k$. An element of $X^{[k]}$ is written as

$$\underline{x} = (x_\epsilon \colon \epsilon \in \{0,1\}^k)\ .$$

We recall that for $\epsilon \in \{0,1\}^k$ and $h \in \mathbb{Z}^k$, we write $|\epsilon| = \epsilon_1 + \dots + \epsilon_k$ and $\epsilon \cdot h = \epsilon_1 h_1 + \dots + \epsilon_k h_k$.

We write the element with all 0's of $\{0,1\}^k$ as $\mathbf{0} = (0, 0, \dots, 0)$. We often give the $\mathbf{0}$-th coordinate of a point of $X^{[k]}$ a distinguished role and we write

$$X^{[k]} = X \times X_*^{[k]}\ ,\ \text{where } X_*^{[k]} = X^{2^k - 1}\ .$$

The coordinates of $X_*^{[k]}$ are indexed by the set

$$\{0,1\}^k_* = \{0,1\}^k \setminus \{\mathbf{0}\}$$

and a point of $X^{[k]}$ is often written

$$\underline{x} = (x_{\mathbf{0}}, \underline{x}_*) \ , \ \text{ where } \underline{x}_* = (x_\epsilon \colon \epsilon \in \{0,1\}^k_*) \ .$$

When (X,μ,T) is a measure preserving system, we also have notation for some transformations that are naturally defined on $X^{[k]}$. Namely, we write $T^{[k]}$ for the transformation $T\times T\times\ldots\times T$, taken 2^k times. Moreover, if $i\in\{1,\ldots,k\}$, we define

$$({T_i}^{[k]}\underline{x})_\epsilon = \begin{cases} T(x_\epsilon) & \text{if } \ \epsilon_i = 1 \\ x_\epsilon & \text{otherwise} \end{cases} .$$

For convenience, we also write $X^{[0]} = X$ and $T^{[0]} = T$.

3.2.2. *Measures and HK-seminorms.* Throughout the rest of this section, (X,μ,T) denotes an ergodic system.

By induction, for every integer $k\geq 0$ we define a measure $\mu^{[k]}$ on $X^{[k]}$ that is invariant under $T^{[k]}$. We set $\mu^{[0]} = \mu$. For $k\geq 1$, making the natural identification of $X^{[k]}$ with $X^{[k-1]}\times X^{[k-1]}$, we write $\underline{x} = (\underline{x}', \underline{x}'')$ for a point of $X^{[k]}$, with $\underline{x}',\underline{x}''\in X^{[k-1]}$. Let $\mathcal{I}^{[k-1]}$ denote the invariant σ-algebra of the system $(X^{[k]},\mu^{[k-1]},T^{[k-1]})$. We define $\mu^{[k]}$ to be the *relatively independent joining* of $\mu^{[k-1]}$ with itself over $\mathcal{I}^{[k-1]}$, meaning that if F,G are bounded functions on $X^{[k-1]}$, then

$$\int_{X^{[k]}} F(\underline{x}')G(\underline{x}'')\,d\mu^{[k]}(\underline{x}) = \int_{X^{[k-1]}} \mathbb{E}(F\mid\mathcal{I}^{[k-1]})(\underline{y})\cdot\mathbb{E}(G\mid\mathcal{I}^{[k-1]})(\underline{y})\,d\mu^{[k-1]}(\underline{y}) \ .$$

By induction, all the marginals of $\mu^{[k]}$ (that is, the images of this measure under the natural projections $X^{[k]}\to X$) are equal to μ.

Since $(X^{[0]},\mu^{[0]},T^{[0]}) = (X,\mu,T)$ is ergodic, $\mathcal{I}^{[0]}$ is the trivial σ-algebra and $\mu^{[1]} = \mu\times\mu$. But for $k\geq 2$ the system $(X^{[k-1]},\mu^{[k-1]},T^{[k-1]})$ is not necessarily ergodic and $\mu^{[k]}$ is not in general the product measure.

For $k\geq 1$ and every $f\in L^\infty(\mu)$,

$$\begin{aligned}\int_{X^{[k]}} \prod_{\epsilon\in\{0,1\}^k} C^{|\epsilon|} f(x_\epsilon)\,d\mu^{[k]}(\underline{x}) \\ = \int_{X^{[k-1]}} \Big|\mathbb{E}\Big(\prod_{\eta\in\{0,1\}^{k-1}} C^{|\eta|} f(y_\eta)\Big|\mathcal{I}^{[k-1]}\Big)\Big|^2\,d\mu^{[k-1]}(\underline{y})\geq 0\end{aligned}$$

and so we can define the *HK-seminorm*

$$|\!|\!| f|\!|\!|_k = \Big(\int_{X^{[k]}}\prod_{\epsilon\in\{0,1\}^k} C^{|\epsilon|} f(x_\epsilon)\,d\mu^{[k]}(\underline{x})\Big)^{1/2^k} \ .$$

To avoid ambiguities when several measures are present, we sometimes write $|\!|\!| f|\!|\!|_{\mu,k}$ instead of $|\!|\!| f|\!|\!|_k$.

In [HK1], we show that $|\!|\!|\cdot|\!|\!|_k$ is a seminorm on $L^\infty(\mu)$. These seminorms satisfy an inequality similar to the Cauchy-Schwartz-Gowers inequality for Gowers norms.

Namely, let f_ϵ, $\epsilon \in \{0,1\}^k$, be 2^k bounded functions on X. Then

$$\Big|\int \prod_{\epsilon\in\{0,1\}^k} f_\epsilon(x_\epsilon)\,d\mu^{[k]}(\underline{x})\Big| \leq \prod_{\epsilon\in\{0,1\}^k} \|\!| f_\epsilon |\!\|_k\ . \tag{2}$$

We also have that consecutive HK-seminorms satisfy $\|\!| f|\!\|_{k+1} \geq \|\!| f|\!\|_k$, and by an application of the ergodic theorem,

$$\|\!| f|\!\|_{k+1} = \lim_{H\to+\infty}\Big(\frac{1}{H}\sum_{h=0}^{H-1} \|\!| T^h f\cdot f|\!\|_k^{2^k}\Big)^{1/2^{k+1}}\ . \tag{3}$$

Using the definition and the fact that the marginals of $\mu^{[k]}$ are equal to μ, we have that for all $f\in L^{2^k}(\mu)$,

$$\|\!| f|\!\|_k \leq \|f\|_{L^{2^k}(\mu)}\ . \tag{4}$$

In fact, the definition of the seminorm $\|\!|\cdot|\!\|_k$ can be extended to $L^{2^k}(\mu)$ with the same properties.

3.3. **Convergence results.**

3.3.1. *Averaging along parallelepipeds.* These seminorms and a geometric description of the factors they define are used to show:

Theorem 3.7 ([HK1], Theorem 13.1)**.** *Let f_ϵ, $\epsilon\in\{0,1\}^k_*$ be 2^k-1 functions in $L^\infty(\mu)$. Then the averages*

$$\frac{1}{H^k}\sum_{h_1,\dots,h_k=0}^{H-1}\ \prod_{\epsilon\in\{0,1\}^k_*} T^{\epsilon\cdot h} f_\epsilon$$

converge in $L^2(\mu)$ and the limit g of these averages is characterized by

$$\int h\,g\,d\mu = \int h(x_{\mathbf{0}}) \prod_{\epsilon\in\{0,1\}^k_*} f_\epsilon(x_\epsilon)\,d\mu^{[k]}(\underline{x})$$

for every $h\in L^\infty(\mu)$.

In fact, we could replace the averages on $[0,H-1]^k$ by averages over any Følner sequence in $\mathbb{Z}^k$. Applying Theorem 3.7 to the case that $f_\epsilon = C^{|\epsilon|}f$ for every ϵ, we obtain:

Corollary 3.8. *For every $f\in L^\infty(\mu)$, the averages*

$$\frac{1}{H^k}\sum_{h_1,\dots,h_k=0}^{H-1}\ \prod_{\epsilon\in\{0,1\}^k_*} C^{|\epsilon|} f(T^{\epsilon\cdot h}x) \tag{5}$$

converge in $L^2(\mu)$ as $H\to+\infty$.

This leads us to a definition:

Definition 3.9. We denote the limit of (5) by $\mathcal{D}_k f$ and call this function the *dual function of f*.

It follows that the dual function $\mathcal{D}_k f$ satisfies:

$$\int \mathcal{D}_k f \, . \, h \, d\mu = \int h(x_{\mathbf{0}}) \prod_{\epsilon\in\{0,1\}^k_*} C^{|\epsilon|} f(x_\epsilon) \, d\mu^{[k]}(\underline{x}) \tag{6}$$

for every $h \in L^\infty(\mu)$. In particular, we have

$$|\!|\!| f |\!|\!|_k^{2^k} = \int \mathcal{D}_k f \, . \, f \, d\mu = \lim_{H\to+\infty} \frac{1}{H^k} \sum_{h_1,\dots,h_k=0}^{H-1} \int \prod_{\epsilon\in\{0,1\}^k} C^{|\epsilon|} T^{\epsilon\cdot h} f \, d\mu \ . \tag{7}$$

The notion of a dual function is implicit in [HK1] and this notation is not used there. However, the notation is coherent with that used in several papers of Green and Tao, where similar functions (in the finite setting) are called dual functions.

The definition extends to functions in $L^{2^k}(\mu)$, for which we use the same notation. Indeed, by (2), (4), and density, for $f \in L^{2^k}(\mu)$ the convergence (5) holds in $L^{2^k/(2^k-1)}(\mu)$; the limit function $\mathcal{D}_k f$ belongs to $L^{2^k/(2^k-1)}(\mu)$ with

$$\|\mathcal{D}_k f\|_{L^{2^k/(2^k-1)}(\mu)} \le \|f\|_{L^{2^k}(\mu)}^{2^k-1}$$

and formula (6) holds for every $h \in L^{2^k}(\mu)$. Moreover, $\mathcal{D}_k$ is a continuous map from $L^{2^k}(\mu)$ to $L^{2^k/(2^k-1)}(\mu)$.

3.3.2. *Application to sequences.* Let f be a bounded function on X. We consider the quantities associated to the bounded sequence $(f(T^n x) \colon n \in \mathbb{Z})$ for a generic point x of X, as in Section 2.1. ¿From the definition of the ergodic seminorms, the pointwise ergodic theorem, and (7), we immediately deduce:

Corollary 3.10. *Let $k \ge 2$ be an integer and let $\mathbf{I}$ be the sequence of intervals $([0, N-1] \colon N \ge 1)$. Let (X, μ, T) be an ergodic system and let $f \in L^\infty(\mu)$. Then for almost every $x \in X$, the sequence $(f(T^n x) \colon n \in \mathbb{Z})$ satisfies property $\mathcal{P}(k)$ on $\mathbf{I}$ and*

$$\|(f(T^n x) \colon n \in \mathbb{Z})\|_{\mathbf{I},k} = |\!|\!| f |\!|\!|_k \ . \tag{8}$$

Corollary 3.11. *Let $k \ge 2$ be an integer, let (X, T) be a uniquely ergodic system with invariant measure μ, and let f be a Riemann integrable function on X. Then for every $x \in X$ and every sequence of intervals $\mathbf{I}$ whose lengths tend to infinity, the sequence $(f(T^n x) \colon n \in \mathbb{Z})$ satisfies property $\mathcal{P}(k)$ on $\mathbf{I}$ and equality (8) holds.*

In particular, for every $x \in X$,

$$\|(f(T^n x) \colon n \in \mathbb{Z})\|_{U(k)} = |\!|\!| f |\!|\!|_k \ .$$

Proof. The hypothesis means that for every $\delta > 0$ there exists two continuous functions g, g' on X with $g \le f \le g'$ and $\int (g' - g) \, d\mu < \delta$. This implies that for every $h \in \mathbb{Z}^k$ the function in the last integral of formula (7) is also Riemann integrable. Therefore the ergodic averages of this function converge everywhere to its integral. □

3.4. **The structure Theorem.** We use the following version of the Structure Theorem of [HK1], which is a combination of statements in Lemma 4.3, Definition 4.10 and Theorem 10.1 of that paper.

Theorem (Structure Theorem)**.** *Let (X, μ, T) be an ergodic system. Then for every $k \ge 2$ there exists a system (Z_k, μ_k, T) and a factor map $\pi_k \colon X \to Z_k$ with the following properties:*

(i) (Z_k, μ_k, T) *is the inverse limit of a sequence of* $(k-1)$*-step nilsystems.*
(ii) *For every function* $f \in L^\infty(\mu)$, $|||f - \mathbb{E}(f \mid Z_k) \circ \pi_k|||_k = 0$.

Since $|||f|||_{k+1} \geq |||f|||_k$ for every $f \in L^\infty(\mu)$, the factors Z_k are nested: Z_k is a factor of Z_{k+1}.

We use this theorem via the following immediate corollary.

Corollary 3.12. *Let* (X, μ, T) *be an ergodic system and* $f \in L^\infty(\mu)$*. Then for every* $\delta > 0$*, there exists a* $(k-1)$*-step ergodic nilsystem* (Y, S, ν)*, a (measure theoretic) factor map* $p\colon X \to Y$*, and a continuous function* h *on* Y *with* $|||f - h \circ p|||_{\mu,k} < \delta$*.*

4. The correspondence principle and the "seminorms"

4.1. **The classic Correspondence Principle.** In translating Szemerédi's Theorem into a problem in ergodic theory, Furstenberg introduced the Correspondence Principle in [F]. We give a not completely classical presentation of this principle, which is amenable to modification in the sequel.

By a *separable subalgebra of* $\ell^\infty(\mathbb{Z})$, we mean a unitary subalgebra of $\ell^\infty(\mathbb{Z})$, invariant under the shift and under complex conjugation, closed in $\ell^\infty(\mathbb{Z})$ and separable for the uniform norm written $\|\cdot\|_\infty$. In the sequel, we mostly consider the case of the separable subalgebra $\mathcal{A}(\mathbf{a})$ spanned by a bounded sequence $\mathbf{a} = (a_n\colon n \in \mathbb{Z})$.

We write σ for the shift on $\ell^\infty(\mathbb{Z})$, and thus for a sequence $\mathbf{a} = (a_n\colon n \in \mathbb{Z})$, $\sigma\mathbf{a}$ denotes the sequence $(a_{n+1}\colon n \in \mathbb{Z})$. We use $\overline{\mathbf{a}}$ to denote the conjugate sequence $(\overline{a}\colon n \in \mathbb{Z})$. In the sequel, $\mathcal{A}$ denotes a separable subalgebra of $\ell^\infty(\mathbb{Z})$.

4.1.1. *The pointed dynamical system associated to an algebra.* Let X be the *Gelfand spectrum* of $\mathcal{A}$, meaning X consists of the set of unitary homomorphisms from $\mathcal{A}$ to the complex numbers. Letting $\mathcal{C}(X)$ denote the algebra of continuous functions on X, we have that there exists an isometric isomorphism of algebras $\Phi\colon \mathcal{C}(X) \to \mathcal{A}$. For $\mathbf{b} \in \mathcal{A}$, the function $\Phi^{-1}(\mathbf{b})$ is called *the function associated to* $\mathbf{b}$.

Since $\mathcal{A}$ is separable, X is a compact metric space. We write d_X for a distance on X defining its topology.

The map $\mathbf{b} \mapsto b_0$ is a character of the algebra $\mathcal{A}$. Thus there exists a point $x_0 \in X$ with $f(x_0) = \Phi(f)_0$ for all $f \in \mathcal{C}(X)$. The shift on $\mathcal{A}$ induces a homeomorphism $T\colon X \to X$ with $\Phi(f \circ T) = \Phi(f) \circ \sigma$ for all $f \in \mathcal{C}(X)$. Therefore, for every $f \in \mathcal{C}(X)$, $\Phi(f)$ is the sequence

$$\Phi(f) = \bigl(f(T^n x_0)\colon n \in \mathbb{Z}\bigr) .$$

In particular, if $f \in \mathcal{C}(X)$ satisfies $f(T^n x_0) = 0$ for all $n \in \mathbb{Z}$, then the sequence given by $\Phi(\mathbf{b}) = f$ is identically zero and so f itself is identically zero. It follows that the point x_0 of X is *transitive*, meaning that its orbit $\{T^n x_0\colon n \in \mathbb{Z}\}$ is dense in X.

We encapsulate this construction in the following definition:

Definition 4.1. The triple (X, T, x_0) is called the *pointed topological dynamical system associated to the algebra* $\mathcal{A}$.

4.1.2. *Averaging schemes and invariant measures.* We first introduce a definition that allows us to average any sequence in a subalgebra over a sequence of intervals:

Definition 4.2. Let $\mathcal{A}$ be a separable subalgebra of $\ell^\infty(\mathbb{Z})$ and $\mathbf{I} = (I_j : j \geq 1)$ be a sequence of intervals whose lengths tend to infinity. We say that $\mathbf{I}$ is an *averaging scheme* for $\mathcal{A}$ if the limit

$$\text{lim averages}_{\mathbf{I}}\big(\mathbf{b}\big) := \lim_{j\to+\infty} \frac{1}{|I_j|} \sum_{n\in I_j} b_n$$

exists for all $\mathbf{b} \in \mathcal{A}$.

Since $\mathcal{A}$ is separable with respect to the norm of $\ell^\infty(\mathbb{Z})$, for every sequence of intervals whose lengths tend to infinity, we can always pass to a subsequence that is an averaging scheme for $\mathcal{A}$. The classical case is when $\mathbf{I}$ is taken to be the sequence $([0, j-1] : j \geq 1)$, or some subsequence of this sequence.

Given an averaging scheme $\mathbf{I}$ for $\mathcal{A}$, we can associate an invariant probability measure μ on X defined by:

$$\int f\, d\mu = \text{lim averages}_{\mathbf{I}}\big(f(T^n x_0)\big) := \lim_{j\to+\infty} \frac{1}{|I_j|} \sum_{n\in I_j} f(T^n x_0) \tag{9}$$

for all $f \in \mathcal{C}(X)$.

We claim that all ergodic invariant probability measures on X are obtained by this procedure. Namely, let μ be such a measure. Let $x_1 \in X$ be a *generic point* for μ, meaning that for all $f \in \mathcal{C}(X)$,

$$\lim_{j\to+\infty} \frac{1}{j} \sum_{n=0}^{j-1} f(T^n x_1) = \int f\, d\mu\ .$$

(By the ergodic theorem, μ-almost every point $x_1 \in X$ is generic.) Since x_0 is a transitive point, there exists a sequence $(k_j : j \geq 1)$ of integers such that

$$\sup_{0\leq n<j} d_X(T^{k_j+n}x_0, T^n x_1) \to 0 \text{ as } j \to +\infty\ .$$

So for any continuous function f on X, we then have

$$\lim_{j\to+\infty} \Big(\frac{1}{j} \sum_{n=0}^{j-1} f(T^n x_1) - \frac{1}{j} \sum_{n=0}^{j-1} f(T^{k_j+n} x_0) \Big) = 0\ .$$

Let $\mathbf{I}$ be the sequence of intervals $(I_j = [k_j, k_j + j - 1] : j \geq 1)$. If $\mathbf{b} \in \mathcal{A}$ and f is the associated function on X, we have

$$\text{lim averages}_{\mathbf{I}}\big(b_n\big) = \text{lim averages}_{\mathbf{I}}\big(f(T^n x_0)\big) = \int f\, d\mu\ .$$

Therefore the sequence of intervals $\mathbf{I}$ is an averaging scheme for $\mathcal{A}$ corresponding to the measure μ, and the claim follows.

4.2. **Proofs of properties of the "seminorms".** We use this presentation of the Correspondence Principle to derive the properties of the "seminorms." We start with the non-negativity that makes the definition possible. Recall that the bounded sequence $\mathbf{a} = (a_n : n \in \mathbb{Z})$ satisfies property $\mathcal{P}(k)$ on the sequence of intervals $\mathbf{I}$ if for all $h = (h_1, \ldots, h_k) \in \mathbb{Z}^k$, the limit

$$c_h(\mathbf{I}, \mathbf{a}) = \text{lim averages}_{\mathbf{I}}\big(\prod_{\epsilon\in\{0,1\}^k} C^{|\epsilon|} a_{n+\epsilon\cdot h}\big)$$

exists. We show that for a sequence $\mathbf{a}$ satisfying this, the limit

$$\lim_{H\to+\infty}\frac{1}{H^k}\sum_{h_1,\dots,h_k=0}^{H-1}c_h(\mathbf{I},\mathbf{a})$$

exists and is non-negative:

Proof of Proposition 2.2. Let $k\geq 2$ be an integer and $\mathbf{a}=(a_n\colon n\in\mathbb{Z})$ be a bounded sequence that satisfies property $\mathcal{P}(k)$ on a sequence of intervals $\mathbf{I}$. Let $\mathcal{A}=\mathcal{A}(\mathbf{a})$, (X,T,x_0) be the pointed topological dynamical system associated to the algebra $\mathcal{A}$, and $f\in\mathcal{C}(X)$ be the function associated to the sequence $\mathbf{a}$. Starting with the sequence of intervals $\mathbf{I}$, by passing to a subsequence $\mathbf{J}$, we extract an averaging scheme for $\mathcal{A}$. Let μ be the associated measure on X. For every $h\in\mathbb{Z}^k$, we have

$$c_h(\mathbf{I},\mathbf{a})=c_h(\mathbf{J},\mathbf{a})=\int\prod_{\epsilon\in\{0,1\}^k}C^{|\epsilon|}f(T^{\epsilon\cdot h}x)\,d\mu(x)\ . \tag{10}$$

Let

$$\mu=\int_\Omega\mu_\omega\,dP(\omega)$$

be the ergodic decomposition of the measure μ. The integral (10) can be rewritten as

$$\int\Bigl(\int\prod_{\epsilon\in\{0,1\}^k}C^{|\epsilon|}f(T^{\epsilon\cdot h}x)\,d\mu_\omega(x)\Bigr)\,dP(\omega)\ .$$

By Theorem 3.7,

$$\lim_{H\to+\infty}\frac{1}{H^k}\sum_{h_1,\dots,h_k=0}^{H-1}c_h(\mathbf{a},\mathbf{I})=\int|\!|\!|f|\!|\!|_{\mu_\omega,k}^{2^k}\,dP(\omega)\ .$$

Therefore, the announced limit exists and is non-negative and we have the statement. □

Maintaining notation used in the proof, we note that:

$$\|\mathbf{a}\|_{\mathbf{I},k}=\Bigl(\int|\!|\!|f|\!|\!|_{\mu_\omega,k}^{2^k}\,dP(\omega)\Bigr)^{1/2^k}\ . \tag{11}$$

We now prove the versions of subadditivity that are satisfied by the "seminorms":

Proof of Propositions 2.4 and 2.5. Assume that the bounded sequence $\mathbf{a}$ satisfies properties $\mathcal{P}(k)$ and $\mathcal{P}(k+1)$ on the sequence of intervals $\mathbf{I}$. By (11), the Cauchy-Schwartz inequality, and equality (3), we have

$$\|\mathbf{a}\|_{\mathbf{I},k}^{2^{k+1}}\leq\int|\!|\!|f|\!|\!|_{\mu_\omega,k}^{2^{k+1}}\,dP(\omega)\leq\int|\!|\!|f|\!|\!|_{\mu_\omega,k+1}^{2^{k+1}}\,dP(\omega)=\|\mathbf{a}\|_{\mathbf{I},k+1}^{2^{k+1}}\ .$$

Thus $\|\mathbf{a}\|_{\mathbf{I},k}\leq\|\mathbf{a}\|_{\mathbf{I},k+1}$ and Proposition 2.5 follows.

Now assume that $\mathbf{a}$ and $\mathbf{b}$ are bounded sequences and assume that the three sequences $\mathbf{a},\mathbf{b}$, and $\mathbf{a}+\mathbf{b}$ satisfy property $\mathcal{P}(k)$ for some sequence of intervals $\mathbf{I}$. We proceed as in the proof of Proposition 2.2, taking $\mathcal{A}$ to be the algebra spanned by $\mathbf{a}$ and $\mathbf{b}$. If f and g are the functions on X associated respectively to $\mathbf{a}$ and $\mathbf{b}$, we have that

$$\|\mathbf{a}\|_{\mathbf{I},k}^{2^k} = \int |\!|\!| f |\!|\!|_{\mu_\omega,k}^{2^k}\, dP(\omega)\ ;\ \|\mathbf{b}\|_{\mathbf{I},k}^{2^k} = \int |\!|\!| g |\!|\!|_{\mu_\omega,k}^{2^k}\, dP(\omega)\ ;$$

$$\|\mathbf{a}+\mathbf{b}\|_{\mathbf{I},k}^{2^k} = \int |\!|\!| f+g |\!|\!|_{\mu_\omega,k}^{2^k}\, dP(\omega)\ .$$

Therefore

$$\|\mathbf{a}+\mathbf{b}\|_{\mathbf{I},k} \leq \|\mathbf{a}\|_{\mathbf{I},k} + \|\mathbf{b}\|_{\mathbf{I},k}$$

and Proposition 2.4 follows. □

4.3. **A Cauchy-Schwartz-Gowers type result.** We have an inequality similar to that satisfied by the Gowers norms in the finite setting and by the HK-seminorms, as given in (2):

Proposition 4.3. *For every $\epsilon \in \{0,1\}^k$, let $\mathbf{a}(\epsilon) = \big(a_n(\epsilon) \colon n \in \mathbb{Z}\big)$ be a bounded sequence. Let $\mathbf{I}$ be a sequence of intervals whose lengths tend to infinity such that*

$$c_h := \lim_{j\to+\infty} \frac{1}{|I_j|} \sum_{n\in I_j} \prod_{\epsilon\in\{0,1\}^k} a_{n+\epsilon\cdot h}$$

exists for every $h \in \mathbb{Z}^k$. Then the limit

$$\lim_{H\to+\infty} \frac{1}{H^k} \sum_{h_1,\dots,h_k=0}^{H-1} c_h$$

exists.

Moreover, if all the sequences $\mathbf{a}(\epsilon)$ satisfy property $\mathcal{P}(k)$ on $\mathbf{I}$, then

$$\Big|\lim_{H\to+\infty} \frac{1}{H^k} \sum_{h_1,\dots,h_k=0}^{H-1} c_h\Big| \leq \prod_{\epsilon\in\{0,1\}^k} \|\mathbf{a}(\epsilon)\|_{\mathbf{I},k}\ . \tag{12}$$

Proof. The proof of the convergence is similar to the proof of Proposition 2.2, but we set $\mathcal{A}$ to be the algebra spanned by the 2^k sequences $\mathbf{a}(\epsilon)$, $\epsilon \in \{0,1\}^k$. Maintaining notation as that proof, for every $\epsilon \in \{0,1\}^k$ we let f_ϵ denote the function associated to the sequence $\mathbf{a}(\epsilon)$. It follows from inequality (2) that

$$\Big|\lim_{H\to+\infty} \frac{1}{H^k} \sum_{h_1,\dots,h_k=0}^{H-1} c_h\Big| = \Big|\int\Big(\int \prod_{\epsilon\in\{0,1\}^k} f_\epsilon(x_\epsilon)\, d\mu_\omega^{[k]}(\underline{x})\Big)\, dP(\omega)\Big|$$

$$\leq \int \prod_{\epsilon\in\{0,1\}^k} |\!|\!| f_\epsilon |\!|\!|_{\mu_\omega,k}\, dP(\omega) \leq \prod_{\epsilon\in\{0,1\}^k} \Big(\int |\!|\!| f_\epsilon |\!|\!|_{\mu_\omega,k}^{2^k}\, dP(\omega)\Big)^{1/2^k} = \prod_{\epsilon\in\{0,1\}^k} \|\mathbf{a}(\epsilon)\|_{\mathbf{I},k}\ .$$

□

Using relations (3) and (11), we deduce that:

Proposition 4.4. *Assume that the bounded sequence $\mathbf{a}$ satisfies property $\mathcal{P}(k+1)$ on $\mathbf{I}$. Then*

$$\lim_{H\to+\infty} \frac{1}{H} \sum_{h=0}^{H-1} \|\sigma^h \mathbf{a}.\bar{\mathbf{a}}\|_{\mathbf{I},k}^{2^k} = \|\mathbf{a}\|_{\mathbf{I},k+1}^{2^{k+1}}\ .$$

Note that the hypothesis implies that for every integer $h \geq 1$, the sequence $\sigma^h \mathbf{a}.\bar{\mathbf{a}}$ satisfies property $\mathcal{P}(k)$ on $\mathbf{I}$.

4.4. **The uniformity seminorms.** We also use the Correspondence Principle to derive properties of the uniformity seminorms:

Proposition 4.5. *Let $k \geq 1$ be an integer,* $\mathbf{a}$ *be a bounded sequence, (X, T, x_0) the associated pointed dynamical system, and $f \in \mathcal{C}(X)$ be the function associated to* $\mathbf{a}$. *Then*

$$\|\mathbf{a}\|_{U(k)} = \sup_{\mu \text{ ergodic}} \|\!|f|\!\|_{\mu,k} \ ,$$

where the supremum is taken over all ergodic measures μ on X.

Proof. It follows from (11) that if we raise the left hand side to the power 2^k, then it is bounded by the right hand side raised to the power 2^k. Conversely, in Section 4.1.2 we showed that every ergodic measure μ on X is associated to an averaging scheme $\mathbf{I}$ for the algebra $\mathcal{A}(\mathbf{a})$. By applying (11) again, we have that $\|\!|f|\!\|_{\mu,k} = \|\mathbf{a}\|_{\mathbf{I},k} \leq \|\mathbf{a}\|_{U(k)}$. □

Proposition 2.7 follows immediately; it could also be derived directly from Proposition 2.4.

Remark 4.1. We note that there are important differences between the uniformity seminorms and the HK-seminorms. For example, the formula given by Proposition 4.4 comes from, and is similar to, formula (3) for the HK-seminorms. We deduce that

$$\|\mathbf{a}\|_{U(k+1)}^{2^{k+1}} \leq \liminf_{H\to+\infty} \frac{1}{H}\sum_{h=0}^{H-1} \|\bar{\mathbf{a}}.\sigma^h\mathbf{a}\|_{U(k)}^{2^k} \ .$$

But in general, the lim inf on the right hand side of this equation is not a limit and equality does not hold.

5. A duality in nilmanifolds and direct results

5.1. **Measures and norms for nilsystems.** Throughout this section, we assume that $k \geq 2$ is an integer and $(X = G/\Gamma, \mu, T)$ is an ergodic $(k-1)$-step nilsystem, where T is the translation by $\tau \in G$. As explained in Section 3, we reduce to the case that G is spanned by its connected component G_0 of the identity and by τ.

We review properties of the measure $\mu^{[k]}$ and of the seminorm $\|\!|\cdot|\!\|_k$ in this particular case. Most of these properties are established in [HK1] or [GT2], but often in a very different context and with very different terminology from that used here. We include some proofs for completeness, but as they are far from the main topics of the article, we defer them to Appendix B. This appendix also includes some properties we need that are not stated elsewhere.

We use the notation for 2^k-Cartesian powers introduced in Section 3. We summarize the properties that we need:

Theorem 5.1.

(i) *The measure $\mu^{[k]}$ is the Haar measure of a sub-nilmanifold X_k of $X^{[k]}$. The transformations $T^{[k]}$ and $T_i^{[k]}$, $1 \leq i \leq k$, act on X_k by translation and X_k is ergodic (and thus uniquely ergodic and minimal) under these transformations.*

(ii) *Let X_{k*} be the image of X_k under the projection $\underline{x} \mapsto \underline{x}_*$ from $X^{[k]}$ to $X_*^{[k]} = X^{2^k-1}$. There exists a smooth map $\Phi\colon X_{k*} \to X_k$ such that*

$$X_k = \left\{(\Phi(\underline{x}_*), \underline{x}_*) \colon \underline{x} \in X_{k*}\right\} \ .$$

(iii) $\|\!|\cdot|\!\|_k$ *is a norm on* $\mathcal{C}(X)$.

(iv) *For every* $x\in X$, *let* $W_{k,x}=\{\underline{x}\in X_k\colon x_{\mathbf{0}}=x\}$. *Then* $W_{k,x}$ *is uniquely ergodic under the transformations* $T_i^{[k]}$, $1\leq i\leq k$.

(v) *For every* $x\in X$, *let* ρ_x *be the invariant measure of* $W_{k,x}$. *Then for every* $x\in X$ *and* $g\in G$, $\rho_{g.x}$ *is the image of* ρ_x *under the translation by* $g^{[k]}=(g,g,\dots,g)$.

The nilmanifold X_k is defined independently of the transformation T and it only depends on the structure of the nilmanifold X. This implies that the measure $\mu^{[k]}$ and the norm $\|\!|\cdot|\!\|_k$ do not depend on the transformation T on X, provided that T is an ergodic transformation. These are geometric, and not dynamical, objects.

5.2. **Uniform convergence.** Using part (iv) of Theorem 5.1 we deduce:

Corollary 5.2. *Let* f_ϵ, $\epsilon\in\{0,1\}^k_*$ *be* 2^k-1 *continuous functions on* X. *For every* $x\in X$ *we have*

$$\frac{1}{H^k}\sum_{h_1,\dots,h_k=0}^{H-1}\prod_{\epsilon\in\{0,1\}^k_*}f_\epsilon(T^{\epsilon\cdot h}x)\to\int\prod_{\epsilon\in\{0,1\}^k_*}f_\epsilon(x_\epsilon)\,d\rho_x(\underline{x})$$

as $H\to+\infty$. *Moreover, the convergence is uniform in* $x\in X$.

Proof. The corollary follows easily from part (iv) of Theorem 5.1 by a classical argument. Let $(x_j\colon j\geq 1)$ be a sequence in X converging to some $x\in X$ and let $(H_j\colon j\geq 1)$ be a sequence of integers tending to infinity.

For every j, let ν_j be the measure

$$\nu_j:=\frac{1}{H_j^k}\sum_{h_1,\dots,h_k=0}^{H_j-1}\bigotimes_{\epsilon\in\{0,1\}^k}\delta_{T^{\epsilon\cdot h}x_j}$$

on $X^{[k]}$ and let ν be any weak limit of this sequence of measures. For every j, the measure ν_j is concentrated on W_{k,x_j}. Since X_k is closed in $X^{[k]}$, the measure ν is concentrated on $W_{k,x}$. Moreover, for every j and for $1\leq i\leq k$, the difference between the measures ν_j and $T_i^{[k]}\nu_j$ are at a distance $\leq 2/H_j$ in the norm of total variation. It follows that ν is invariant under $T_i^{[k]}$ for $i=1,\dots,k$. By unique ergodicity of $W_{k,x}$, we have that ν is the invariant measure ρ_x of $W_{k,x}$.

We have shown that the sequence $(\nu_j\colon j\geq 1)$ of measures converges weakly to the measure ρ_x. It follows that if f_ϵ, $\epsilon\in\{0,1\}^k_*$, are continuous functions on X, then

$$\begin{aligned}\frac{1}{H_j^k}\sum_{h_1,\dots,h_k=0}^{H_j-1}\prod_{\epsilon\in\{0,1\}^k_*}f_\epsilon(T^{\epsilon\cdot h}x_j)&=\int\prod_{\epsilon\in\{0,1\}^k_*}f_\epsilon(x_\epsilon)\,d\nu_j(\underline{x})\\&\to\int\prod_{\epsilon\in\{0,1\}^k_*}f_\epsilon(x_\epsilon)\,d\rho_x(\underline{x})\end{aligned}$$

as $j\to+\infty$ and the result follows. □

We apply this result when f is a continuous function on X and $f_\epsilon=C^{|\epsilon|}f$ for every $\epsilon\in\{0,1\}^k_*$. From Corollary 3.8, we have that the averages in Corollary 5.2 converge in $L^2(\mu)$ to the function $\mathcal{D}_kf$. Therefore:

Corollary 5.3. *Let f be a continuous function on X. Then*

$$\mathcal{D}_k f(x) = \int \prod_{\epsilon\in\{0,1\}^k_*} C^{|\epsilon|} f(x_\epsilon)\, d\rho_x(\underline{x})$$

and the function $\mathcal{D}_k f$ is the uniform limit of the sequence

$$\frac{1}{H^k}\sum_{h_1,\dots,h_k=0}^{H-1}\prod_{\epsilon\in\{0,1\}^k_*} f_\epsilon(T^{\epsilon\cdot h}x)\ .$$

Thus $\mathcal{D}_k f$ is a continuous function on X.

In particular, the function $\mathcal{D}_k f$ is a geometric object: it does not depend on the transformation T on X.

Corollary 5.4. *If f is a smooth function on X, then $\mathcal{D}_k f$ is a smooth function on X.*

Proof. Let $x_0 \in X$. Then, by Corollary 5.3 and part (v) of Theorem 5.1, for every $g \in G$ we have

$$\mathcal{D}_k f(g.x_0) = \int \prod_{\epsilon\in\{0,1\}^k_*} C^{|\epsilon|} f(g.x_\epsilon)\, d\rho_{x_0}(\underline{x})\ .$$

Thus the function $g \mapsto \mathcal{D}_k f(g.x_0)$ is a smooth function on G and the result follows. $\square$

Remark 5.1. Let $x \in X$. Since the measure ρ_x is invariant under the transformations $T_i^{[k]}$, it follows that the image of this measure under the projection $\underline{x} \mapsto x_\epsilon$ for every $\epsilon \in \{0,1\}^k$ is invariant under T and thus is equal to μ. Therefore if f_ϵ, $\epsilon \in \{0,1\}^k_*$, are continuous functions on X, the Hölder inequality gives:

$$\Big|\int \prod_{\epsilon\in\{0,1\}^k_*} f_\epsilon(x_\epsilon)\, d\rho_x(\underline{x})\Big| \le \prod_{\epsilon\in\{0,1\}^k_*} \|f_\epsilon\|_{L^{2^k-1}(\mu)}\ .$$

By density we deduce that for every $f \in L^{2^k-1}(\mu)$ the function $\mathcal{D}_k f$ is continuous on X and that

$$\|\mathcal{D}_k f\|_\infty \le \|f\|^{2^k-1}_{L^{2^k-1}(\mu)}\ .$$

5.3. **The dual norm.**

Definition 5.5. Let the space $\mathcal{C}(X)$ of continuous functions on X be endowed with the norm $\|\!|\cdot|\!\|_k$. Since $\|\!|f|\!\|_k \le \|f\|_{L^{2^k}(\mu)}$ for every $f \in \mathcal{C}(X)$, the dual of this space can be identified with a subspace of $L^{2^k/(2^k-1)}(\mu)$. We call this space *the dual space* and denote it by $\mathcal{C}(X)^*_k$. We write $\|\!|h|\!\|^*_k$ for the *dual norm* of a function $h \in \mathcal{C}(X)^*_k$.

In other words, a function $h \in L^{2^k/(2^k-1)}(\mu)$ belongs to the dual space $\mathcal{C}(X)^*_k$ if there exists a constant C with

$$\Big|\int f\, h\, d\mu\Big| \le C\, \|\!|f|\!\|_k \tag{13}$$

for every $f \in \mathcal{C}(X)$ and $\|\!|f|\!\|^*_k$ is the smallest constant C with this property.

We note that the dual space and the dual norm $\|\!|\cdot|\!\|^*_k$ are geometric, not dynamical, objects.

We give two methods to build functions in the dual space. Let f be a function on X, belonging to $L^{2^k}(\mu)$. By characterization (5) of the dual function and inequality (2), we have that for every $h \in \mathcal{C}(X)$,

$$\Big|\int h.\mathcal{D}_k f\,d\mu\Big| \leq |||h|||_k\,|||f|||_k^{2^k-1}\ .$$

Thus $\mathcal{D}_k f$ belongs to the dual space and $|||\mathcal{D}_k f|||_k^* \leq |||f|||_k^{2^k-1}$. On the other hand,

$$|||f|||_k\,|||\mathcal{D}_k f|||_k^* \geq \int f\,\mathcal{D}_k f\,d\mu = |||f|||_k^{2^k}$$

and we conclude that

$$|||\mathcal{D}_k f|||_k^* = |||f|||_k^{2^k-1}\ . \tag{14}$$

We now show:

Proposition 5.6. *The dual space $\mathcal{C}(X)_k^*$ contains all smooth functions on X.*

Proof. Let f be a smooth function on X and let X_{k*} and Φ be the set and the map defined in part (ii) of Theorem 5.1.

Then $f\circ\Phi$ is a smooth function on X_{k*} and there exists a smooth function F on $X_*^{[k]}$ whose restriction to X_{k*} is equal to $f\circ\Phi$. This function can be written as

$$F(\underline{x}_*) = \sum_{j=1}^{\infty}\prod_{\epsilon\in\{0,1\}_*^k} f_{j,\epsilon}(x_\epsilon)\ ,$$

where the functions $f_{j,\epsilon}$, $j\geq 1$ and $\epsilon\in\{0,1\}_*^k$, are continuous functions on X satisfying

$$\sum_{j=1}^{\infty}\prod_{\epsilon\in\{0,1\}_*^k}\|f_{j,\epsilon}\|_\infty < +\infty\ .$$

For every continuous function h on X, we have

$$\begin{aligned}
\Big|\int f\,h\,d\mu\Big| =&\Big|\int h(x_{\mathbf{0}})\,.\,f\circ\Phi(\underline{x}_*)\,d\mu^{[k]}(\underline{x})\Big| = \Big|\int h(x_{\mathbf{0}})\,.\,F(\underline{x}_*)\,d\mu^{[k]}(\underline{x})\Big|\\
\leq&\sum_{j=1}^{\infty}\Big|\int h(x_{\mathbf{0}})\prod_{\epsilon\in\{0,1\}_*^k} h_{j,\epsilon}(x_\epsilon)\,d\mu^{[k]}(\underline{x})\Big|\\
\leq&\sum_{j=1}^{\infty}|||h|||_k\prod_{\epsilon\in\{0,1\}_*^k}|||h_{j,\epsilon}|||_k\\
\leq&|||h|||_k\sum_{j=1}^{\infty}\prod_{\epsilon\in\{0,1\}_*^k}\|h_{j,\epsilon}\|_\infty\ .
\end{aligned}$$

where the next to last inequality follows from (2). The announced statement follows. $\square$

A similar proof is used in [GT2] in the finite setting.

The hypothesis of smoothness is too strong and could be replaced by weaker assumptions. It is probably sufficient to assume that f is Lipschitz with respect to some smooth metric on X. Computing the dual norm of f, or even bounding it in an explicit way seems to be difficult. The regularity of the map Φ should play a role, but in order to define this, we would first need to choose a metric on X.

Proposition 5.7. *The unit ball of $\mathcal{C}(X)^*_k$ is the closure in $L^{2^k/(2^k-1)}(\mu)$ of the convex hull of the set*
$$\{\mathcal{D}_k f\colon f\in\mathcal{C}(X),\ \| f\|_k\leq 1\}\ .$$

Proof. Let B be the set in the statement. By (14), for $f\in\mathcal{C}(X)$ with $\| f\|_k\leq 1$, we have that $\mathcal{D}_k f$ belongs to the unit ball of $\mathcal{C}(X)^*_k$. Since this ball is closed in the norm of $L^{2^k/(2^k-1)}(\mu)$, it contains B.

On the other hand, let f be a nonzero function belonging to $L^{2^k}(\mu)$ and let $h=\| f\|_k^{-1}.f$. As the map $\mathcal{D}_k\colon L^{2^k}(\mu)\to L^{2^k/(2^k-1)}(\mu)$ is continuous, by density we have that $\mathcal{D}_k h\in B$. As $\int f.\mathcal{D}_k h\,d\mu=\| f\|_k$, the Hahn-Banach Theorem gives the opposite inclusion. □

5.4. **Direct theorem (upper bound).** We now have assembled the ingredients to prove Theorem 2.13. As we have not yet defined the norm $\|\mathbf{b}\|^*_k$ of a smooth nilsequence $\mathbf{b}$, we state this theorem in a modified version.

Theorem (Modified Direct Theorem)**.** *Let $\mathbf{a}=(a_n\colon n\in\mathbb{Z})$ be a bounded sequence that satisfies property $\mathcal{P}(k)$ on the sequence of intervals $\mathbf{I}=(I_j\colon j\geq 1)$. Let (X,T,μ) be an ergodic $(k-1)$-step nilsystem, $x_0\in X$, and f be a smooth function on X. Then*
$$\limsup_{j\to+\infty}\Bigl|\frac{1}{|I_j|}\sum_{n\in I_j}a_n\,f(T^nx_0)\Bigr|\leq\|\mathbf{a}\|_{\mathbf{I},k}\,\| f\|^*_k\ .$$

Proof.

5.4.1. We begin with the case that $f=\mathcal{D}_k\phi$ for some continuous function ϕ on X with $\|\phi\|_k=1$.

By substituting a subsequence for $\mathbf{I}$, we can assume that for every $h=(h_1,\dots,h_k)\in\mathbb{Z}^k$, the averages on I_j of
$$a_n\prod_{\epsilon\in\{0,1\}^k_*}C^{|\epsilon|}\phi(T^{n+\epsilon\cdot h}x_0)$$
converge.

Fix $\delta>0$. By Corollary 5.3, for every sufficiently large H we have that
$$\Bigl|\frac{1}{H^k}\sum_{h_1,\dots,h_k=0}^{H-1}a_n\prod_{\epsilon\in\{0,1\}^k_*}C^{|\epsilon|}\phi(T^{n+\epsilon\cdot h}x_0)-a_nf(T^nx_0)\Bigr|<\delta$$
for every $n\in\mathbb{Z}$ and so
$$\Bigl|\frac{1}{H^k}\sum_{h_1,\dots,h_k=0}^{H-1}\Bigl(\frac{1}{|I_j|}\sum_{n\in I_j}a_n\prod_{\epsilon\in\{0,1\}^k_*}C^{|\epsilon|}\phi(T^{n+\epsilon\cdot h}x_0)\Bigr)-\frac{1}{|I_j|}\sum_{n\in I_j}a_nf(T^nx_0)\Bigr|<\delta\ .$$
for every $j\geq 1$. Taking the limit as $j\to+\infty$ along a subsequence, we have that for every sufficiently large H,
$$\begin{aligned}&\text{limsup}\,|\text{averages}_{\mathbf{I}}(a_nf(T^nx_0))|\\&\qquad\leq\delta+\left|\frac{1}{H^k}\sum_{h_1,\dots,h_k=0}^{H-1}\lim\ \text{averages}_{\mathbf{I}}\bigl(a_n\prod_{\epsilon\in\{0,1\}^k_*}C^{|\epsilon|}\phi(T^{n+\epsilon\cdot h}x_0)\bigr)\right|\ .\end{aligned}$$

We conclude that

$$\begin{aligned}&\operatorname{limsup}|\text{averages}_{\mathbf{I}}(a_nf(T^nx_0))|\\&\qquad\le\Bigl|\lim_{H\to+\infty}\frac{1}{H^k}\sum_{h_1,\dots,h_k=0}^{H-1}\lim\text{averages}_{\mathbf{I}}\bigl(a_n\prod_{\epsilon\in\{0,1\}^k_*}C^{|\epsilon|}\phi(T^{n+\epsilon\cdot h}x_0)\bigr)\Bigr|\ .\end{aligned}$$

The existence of the limit for $H\to+\infty$ is given by Proposition 4.3. Using Inequality (12) and Corollary 3.11, we have that the last quantity is bounded by

$$\|\mathbf{a}\|_{\mathbf{I},k}\,.\,\|(\phi(T^nx_0)\colon n\in\mathbb{Z})\|_{\mathbf{I},k}^{2^k-1}=\|\mathbf{a}\|_{\mathbf{I},k}\,.\,|\!|\!|\phi|\!|\!|_k^{(2^k-1)/2^k}=\|\mathbf{a}\|_{\mathbf{I},k}\ .$$

5.4.2. We now turn to the general case. We can assume that $|\!|\!|f|\!|\!|_k^*\le 1$.

Fix $\delta>0$. By Proposition 5.6, we can write $f=f_1+f_2$, where f_1 is a convex combination of functions considered in the first part and $\|f_2\|_{L^{2^k/(2^k-1)}(\mu)}<\delta$. The contribution of f_1 to the lim sup of the averages is bounded by 1.

For every $j\ge1$, by the Hölder inequality we have

$$\Bigl|\frac{1}{|I_j|}\sum_{n\in I_j}a_nf_2(T^nx_0)\Bigr|\le\|\mathbf{a}\|_\infty\Bigl(\frac{1}{|I_j|}\sum_{n\in I_j}|f_2(T^nx_0)|^{2^k/(2^k-1)}\Bigr)^{(2^k-1)/2^k}\ .$$

Since both f and f_1 are continuous, so is f_2. Therefore, by unique ergodicity of (X,T), the averages of $|f_2(T^nx_0)|^{2^n/(2^n-1)}$ converge to the integral of the function $|f|^{2^n/(2^n-1)}$ and we have that

$$\operatorname{limsup}|\text{averages}_{\mathbf{I}}(a_nf_2(T^nx_0))|\le\delta\ .$$

The result follows. □

5.5. **The dual norm for smooth nilsequences.**

Corollary 5.8. *Let (X,μ,T) and (Y,ν,S) be ergodic $(k-1)$-step nilsystems, $x_0\in X$, $y_0\in Y$, f be a smooth function on X, and g a smooth function on Y. If $f(T^nx_0)=g(S^ny_0)$ for every $n\in\mathbb{Z}$, then $|\!|\!|f|\!|\!|^*_{\mu,k}=|\!|\!|g|\!|\!|^*_{\nu,k}$.*

Proof. Fix $\delta>0$. By definition of $|\!|\!|f|\!|\!|^*_{\mu,k}$, there exists a continuous function h on X with

$$|\!|\!|h|\!|\!|_{\mu,k}=1\text{ and }\Bigl|\int f\,h\,d\mu\Bigr|\ge|\!|\!|f|\!|\!|^*_{\mu,k}-\delta\ .$$

By unique ergodicity of X,

$$\Bigl|\int f\,h\,d\mu\Bigr|=\lim_{N\to+\infty}\Bigl|\sum_{n=0}^{N-1}f(T^nx_0)h(T^nx_0)\Bigr|=\lim_{N\to+\infty}\Bigl|\sum_{n=0}^{N-1}g(S^ny_0)h(T^nx_0)\Bigr|\ .$$

Let $\mathbf{I}$ be the sequence of intervals $(I_N=[0,N-1]\colon N\ge1)$. By Corollary 3.11, the sequence $(h(T^nx_0)\colon n\in\mathbb{Z})$ satisfies property $\mathcal{P}(k)$ on $\mathbf{I}$ and $\|(h(T^nx_0)\colon n\in\mathbb{Z})\|_{\mathbf{I},k}=|\!|\!|h|\!|\!|_{\mu,k}=1$. By the Modified Direct Theorem, we have that

$$\lim_{N\to+\infty}\Bigl|\sum_{n=0}^{N-1}g(S^ny_0)h(T^nx_0)\Bigr|\le|\!|\!|g|\!|\!|^*_{\nu,k}$$

and so $|\!|\!|f|\!|\!|^*_{\mu,k}-\delta\le|\!|\!|g|\!|\!|^*_{\nu,k}$. Exchanging the roles of f and g, we obtain the announced equality. □

Using this corollary, we define:

Definition 5.9. Let $\mathbf{b}$ be a $(k-1)$-step smooth nilsequence. We define $\|\!|\mathbf{b}|\!\|_k^* = \|\!|f|\!\|_{\mu,k}^*$, where f is a smooth function on an ergodic $(k-1)$-step nilsystem (X,μ,T) and $x_0 \in X$ is chosen such that $b_n = f(T^n x_0)$ for every n.

Using this definition, the Direct Theorem (Theorem 2.13) is a reformulation of the Modified Direct Theorem of Section 5.4.

5.6. **The case** $k = 2$**.** Let X be a 1-step nilmanifold, that is, a compact abelian Lie group, and let f be a smooth function on X. Let $\widehat{X}$ be the dual group of G. Then the Fourier series of f is

$$f(x) = \sum_{\chi\in\widehat{X}} \widehat{f}(\chi)\,\chi(x) \ , \text{ where } \sum_{\chi\in\widehat{X}} |\widehat{f}(\chi)| < +\infty \ .$$

An easy computation using the definition gives

$$\|\!|f|\!\|_2 = \Bigl(\sum_{\chi\in\widehat{X}} |\widehat{f}(\chi)|^4\Bigr)^{1/4} \ .$$

Therefore we have

$$\|\!|f|\!\|_2^* = \Bigl(\sum_{\chi\in\widehat{X}} |\widehat{f}(\chi)|^{4/3}\Bigr)^{3/4} \ .$$

If T is an ergodic translation on X, $x_0 \in X$, and $\mathbf{b}$ is the sequence given by $b_n = f(T^n x_0)$ for every n, we recover the formula for $\|\!|\mathbf{b}|\!\|_2^*$ given in Section 2.3 and Proposition 2.12.

5.7. **Some convergence results.**

Corollary 5.10. *Let $k \geq 2$ be an integer, $\mathbf{I} = (I_j : j \geq 1)$ be a sequence of intervals whose lengths tend to infinity, and let $\mathbf{a} = (a_n : n \in \mathbb{Z})$ be a bounded sequence. Assume that for every $\delta > 0$, there exists a $(k-1)$-step nilsequence $\mathbf{a}'$ such that the sequence $\mathbf{a} - \mathbf{a}'$ satisfies property $\mathcal{P}(k)$ on $\mathbf{I}$ and $\|\mathbf{a} - \mathbf{a}'\|_{\mathbf{I},k} < \delta$. Then for every $(k-1)$-step nilsequence $\mathbf{b} = (b_n : n \in \mathbb{Z})$, the limit*

$$\lim_{j\to+\infty} \frac{1}{|I_j|} \sum_{n\in I_j} a_n b_n$$

exists.

Proof. By density, we can restrict to the case that $\mathbf{b}$ is a smooth nilsequence. Let $\delta > 0$ and the nilsequence $\mathbf{a}'$ be as in the statement. Since the product sequence $\mathbf{a}'\mathbf{b}$ is a nilsequence, its averages converge. By Theorem 2.13,

$$\operatorname{limsup} |\text{averages}_{\mathbf{I}}((a_n - a'_n)b_n)| \leq \delta \|\!|b|\!\|_k^* \ .$$

It follows that the averages on I_j of $a_n b_n$ form a Cauchy sequence. □

By the same argument, we have:

Corollary 5.11. *Let $k \geq 2$ be an integer and $\mathbf{a} = (a_n : n \in \mathbb{Z})$ be a bounded sequence. Assume that for every $\delta > 0$, there exists a $(k-1)$-step nilsequence $\mathbf{a}'$ such that $\|\mathbf{a} - \mathbf{a}'\|_{U(k)} < \delta$. Then for every $(k-1)$-step nilsequence $\mathbf{b} = (b_n : n \in \mathbb{Z})$, the averages of the sequence $a_n b_n$ converge, meaning that the limit*

$$\lim_{j\to+\infty} \frac{1}{|I_j|} \sum_{n\in I_j} a_n b_n$$

exists for all sequences of intervals $\mathbf{I} = (I_j : j \geq 1)$ *whose lengths tend to infinity.*

This Corollary is the direct implication of Theorem 2.19. Propositions 2.20 and 2.21 provide examples of sequences satisfying the hypothesis of this Corollary.

6. The correspondence principle revisited and inverse theorems

6.1. **An extension of the correspondence principle.** We recall that a topological dynamical system (Y, S) is *distal* if for every $y, y' \in Y$ with $y \neq y'$, then

$$\inf_{n\in\mathbb{Z}} d_Y(T^n y, T^n y') > 0$$

where d_Y denotes a distance defining the topology of Y.

Proposition 6.1. *Let* (X, T) *be a topological dynamical system,* $x_0 \in X$ *a transitive point, and* μ *an invariant ergodic measure on* X*. Let* (Y, S) *be a distal topological dynamical system,* ν *an invariant measure on* Y*, and* $\pi \colon (X, \mu, T) \to (Y, \nu, S)$ *a measure theoretic factor map.*

Then there exist a point $y_0 \in Y$ *and a sequence of intervals* $\mathbf{I} = (I_j : j \geq 1)$ *whose lengths tend to infinity such that for every continuous function* f *on* X *and every continuous function* g *on* Y*,*

$$\int f(x).g\circ\pi(x)\, d\mu(x) = \lim_{j\to+\infty} \frac{1}{|I_j|} \sum_{n\in I_j} f(T^n x_0).g(S^n y_0)\ .$$

If the system (X, T) and the point x_0 are associated to a sequence as in Section 4.1 and if Y denotes the Kronecker factor of (X, μ, T), then the sequence of intervals $\mathbf{I}$ given by the Proposition plays the same role as the "Kronecker complete processes" of [BFW]. Our construction is (we hope) simpler and works in a more general setting: below we use it when Y is a nilsystem.

Proof. We write $d_X(\cdot,\cdot)$ and $d_Y(\cdot,\cdot)$ for distances on X and Y defining the topologies of these spaces.

6.1.1. *Construction of an extension of* X*.* Let $\mathcal{B}$ be the closed (in norm) subalgebra of $L^\infty(\mu)$ that is spanned by $\mathcal{C}(X)$ and the functions $g \circ \pi$ with $g \in \mathcal{C}(Y)$. This algebra is unitary, separable, and invariant under complex conjugation and under T.

Let W be the Gelfand spectrum of this algebra. Since $\mathcal{B}$ is separable, W is a compact metrizable space. By definition, there exists an isometric isomorphism of algebras $\Psi \colon \mathcal{C}(W) \to \mathcal{B}$.

As in Section 4.1.1, there exists a homeomorphism $R \colon W \to W$ satisfying $\Psi(f \circ T) = \Psi(f) \circ R$ for all functions $f \in \mathcal{C}(W)$.

The inclusion of $\mathcal{C}(X)$ in $\mathcal{B}$ induces a continuous surjective map $p \colon W \to X$ satisfying $f \circ p = \Psi(f)$ for every continuous function f on X and we have that $T \circ p = p \circ R$. Similarly, the map $g \mapsto g \circ \pi$ from $\mathcal{C}(Y)$ to $\mathcal{B}$ is an isometric homomorphism of algebras and thus induces a continuous surjective map $q \colon W \to Y$ satisfying $g \circ q = \Psi(g \circ \pi)$ for all continuous functions g on Y. We have that $S \circ q = q \circ R$. So, $p \colon (W, R) \to (X, T)$ and $q \colon (W, R) \to (Y, S)$ are factor maps, in the topological sense.

The map $f \mapsto \int f\,d\mu$ is a positive linear form on the algebra $\mathcal{B}$ and thus there exists a unique probability measure ρ on W satisfying

$$\int f\,d\mu = \int \Psi(f)\,d\rho \text{ for all functions } f \in \mathcal{B}\ .$$

Since $\Psi(f \circ T) = \Psi(f) \circ R$ for all $f \in \mathcal{B}$ and μ is invariant under T, the measure ρ is invariant under R. Since $\Psi(f) = f \circ p$ for all continuous functions f on X, we have that the image of ρ under p is equal to μ. Therefore, $p\colon (W, \rho, R) \to (X, \mu, T)$ is a measure theoretic factor map. Moreover, for every function $f \in \mathcal{B}$,

$$\int |\Psi(f)|^2\,d\rho = \int \Psi(|f|^2)\,d\rho = \int |f|^2\,d\mu$$

and the map Ψ is an isometry from the space $\mathcal{B}$ endowed with the norm $L^2(\mu)$ into the space $L^2(\rho)$. Since $\mathcal{C}(X)$ is dense in $\mathcal{B}$ under the $L^2(\mu)$ norm and since $\Psi(f) = f \circ p$ for $f \in \mathcal{C}(X)$, we have that for all $f \in \mathcal{B}$,

$$\Psi(f) = f \circ p \quad (\rho\text{-almost everywhere}).$$

We claim that the map $p\colon (W, \rho, R) \to (X, \mu, T)$ is an isomorphism between measure preserving systems. Indeed, the range of the map $f \mapsto f \circ p\colon L^2(\mu) \to L^2(\rho)$ is closed in $L^2(\rho)$ because this map is an isometry, and it contains $\Psi(\mathcal{B}) = \mathcal{C}(W)$ and thus it is equal to $L^2(\rho)$. In particular, (W, ρ, R) is ergodic.

Finally, for every function $g \in \mathcal{C}(Y)$, we have that $g \circ q = \Psi(g \circ \pi) = g \circ \pi \circ p$ (ρ-almost everywhere) and so $q = \pi \circ p$ (ρ-almost everywhere).

In particular, the image of ρ under q is ν.

6.1.2. *Construction of the sequence of intervals.* Since ρ is ergodic under R, it admits a generic point w_1. Recall that this means that for every $f \in \mathcal{C}(W)$,

$$\lim_{j\to+\infty} \frac{1}{j} \sum_{n=0}^{j-1} f(R^n w_1) = \int f\,d\rho\ .$$

Set $x_1 = p(w_1)$. Since x_0 is a transitive point of X, we can choose as in Section 4.1.2 a sequence of integers $(k_j\colon j \geq 1)$ such that

$$\lim_{j\to+\infty} \sup_{0\leq n\leq j} d_X(T^n x_1, T^{k_j+n} x_0) = 0\ . \tag{15}$$

Set $y_1 = q(w_1)$. Let η be a point in the closure of the sequence $(S^{k_j}\colon j \geq 1)$ in the Ellis semigroup [E] of (Y, S). Since (Y, S) is distal, we have (see [A], chapter 5) that η is a bijection from Y onto itself. Pick $y_0 \in Y$ such that $\eta(y_0) = y_1$. Thus passing, if necessary, to a subsequence of $(k_j\colon j \geq 1)$, which we also denote by $(k_j\colon j \geq 1)$, we have that $T^{k_j} y_0$ converges to y_1. Again replacing this sequence by a subsequence, we can assume that

$$\lim_{j\to+\infty} \sup_{0\leq n<j} d_Y(S^n y_1, S^{k_j+n} y_0) = 0\ . \tag{16}$$

For all $j \geq 1$, set $I_j = [k_j, k_j + j - 1]$. Let f be a continuous function on X and g a continuous function on Y. By (15) and (16) we have that

$$\lim_{j\to+\infty} \sup_{0\leq n<j} \bigl|f(T^n x_1) - f(T^{k_j+n} x_0)\bigr| = 0 \text{ and}$$
$$\lim_{j\to+\infty} \sup_{0\leq n<j} \bigl|g(S^n y_1) - g(S^{k_j+n} y_0)\bigr| = 0\ .$$

Thus

$$\lim_{j\to+\infty}\Big(\frac{1}{|I_j|}\sum_{n\in I_j} f(T^nx_0)g(S^ny_0)-\frac{1}{j}\sum_{n=0}^{j-1} f(T^nx_1)g(S^ny_1)\Big)=0\ . \tag{17}$$

For each integer n,

$$f(T^nx_1)g(S^ny_1)=f\circ p(R^nw_1).g\circ q(R^nw_1)\ .$$

Since w_1 is a generic point with respect to the measure ρ, the second average in (17) converges to

$$\int (f\circ p).(g\circ q)\,d\rho=\int (f\circ p).(g\circ\pi\circ p)\,d\rho=\int f.(g\circ\pi)\,d\mu$$

because $q=\pi\circ p$ (ρ-almost everywhere) and the image of ρ under p is μ. □

6.2. **Inverse results.**

Proposition 6.2. *Let $k\geq 2$ be an integer, $\mathbf{a}$ be a bounded sequence, and $\delta>0$. Then there exists a sequence of intervals $\mathbf{I}=(I_j\colon j\geq 1)$ whose lengths tend to infinity and a $(k-1)$-step smooth nilsequence $\mathbf{b}$ such that*

(i) *The sequence $\mathbf{a}$ satisfies property $\mathcal{P}(k)$ on $\mathbf{I}$ and $\|\mathbf{a}\|_{\mathbf{I},k}\geq\|\mathbf{a}\|_{U(k)}-\delta$.*
(ii) *The sequence $\mathbf{a}-\mathbf{b}$ satisfies property $\mathcal{P}(k)$ on $\mathbf{I}$ and $\|\mathbf{a}-\mathbf{b}\|_{\mathbf{I},k}<\delta$.*

Proof. Let (X,T,x_0) be the pointed dynamical system associated to the algebra spanned by the sequence $\mathbf{a}$, as in Section 4.1.1. Let f be the continuous function on X defined by $f(T^nx_0)=a_n$ for every $n\in\mathbb{Z}$.

By Proposition 4.5, there exists an invariant ergodic measure μ on X with $|||f|||_{\mu,k}\geq\|\mathbf{a}\|_{U(k)}-\delta$. By Corollary 3.12 of the Structure Theorem there exist a $(k-1)$-step nilsystem (Y,S,ν), a measure theoretic factor map $\pi\colon(X,\mu,T)\to(Y,\nu,S)$, and a smooth function g on Y with $|||f-g\circ\pi|||_{\mu,k}<\delta$.

Recall that every nilsystem is distal. Now, let $\mathbf{I}$ and y_0 be given by Proposition 6.1 and let $\mathbf{b}$ be the nilsequence given by $b_n=g(S^ny_0)$ for every $n\in\mathbb{Z}$.

The measure on X associated to $\mathbf{I}$ as in 4.1.2 is equal to μ. Thus the sequence $\mathbf{a}$ satisfies property $\mathcal{P}(k)$ on $\mathbf{I}$ and $\|\mathbf{a}\|_{\mathbf{I},k}=|||f|||_{\mu,k}\geq\|\mathbf{a}\|_{U(k)}-\delta$. To prove Proposition 6.2, we are left with proving that the sequence $\mathbf{a}-\mathbf{b}$ satisfies property $\mathcal{P}(k)$ on $\mathbf{I}$ and that $\|\mathbf{a}-\mathbf{b}\|_{\mathbf{I},k}<\delta$.

For $h=(h_1,\dots,h_k)\in\mathbb{Z}^k$, we have

$$\begin{aligned}\prod_{\epsilon\in\{0,1\}^k} C^{|\epsilon|}(a_{n+\epsilon\cdot h}-b_{n+\epsilon\cdot h})&=\prod_{\epsilon\in\{0,1\}^k} C^{|\epsilon|}\big(f(T^{n+\epsilon\cdot h}x_0)-g(S^{n+\epsilon\cdot h}y_0)\big)\\ &=\sum_{(A,B)\text{ partition of }\{0,1\}^k}(-1)^{|B|}\prod_{\epsilon\in A}C^{|\epsilon|}f(T^{n+\epsilon\cdot h}x_0)\prod_{\epsilon\in B}C^{|\epsilon|}g(S^{n+\epsilon\cdot h}y_0)\ .\end{aligned}$$

By definition of $\mathbf{I}$, the averages (with respect to n) of the above expression on this sequence of intervals converge to

$$\begin{aligned}\sum_{(A,B)\text{ partition of }\{0,1\}^k}(-1)^{|B|}&\int\prod_{\epsilon\in A}C^{|\epsilon|}f(T^{\epsilon\cdot h}x)\prod_{\epsilon\in B}C^{|\epsilon|}g\circ\pi(T^{\epsilon\cdot h}x)\,d\mu(x)\\ &=\int\prod_{\epsilon\in\{0,1\}^k}C^{|\epsilon|}(f-g\circ\pi)(T^{\epsilon\cdot h}x)\,d\mu(x)\ .\end{aligned}$$

By definition, the averages (with respect to $h \in \mathbb{Z}^d$) of the first term converge to $\|\mathbf{a}-\mathbf{b}\|_{\mathbf{I},k}$ and, by Corollary 3.8, the averages of the last integral converge to $\|\!|f - g\circ\pi|\!\|_{\mu,k} < \delta$ and we are done. □

We now prove the Inverse Theorem (Theorem 2.16). We recall the statement here for convenience.

Theorem. *Let $\mathbf{a} = (a_n : n \in \mathbb{Z})$ be a bounded sequence. Then for every $\delta > 0$, there exists a $(k-1)$-step smooth nilsequence $\mathbf{b} = (b_n : n \in \mathbb{Z})$ such that*

$$\|\!|\mathbf{b}|\!\|_k^* = 1 \text{ and } \lim_{N\to+\infty} \sup_{M\in\mathbb{Z}} \Big|\frac{1}{N}\sum_{n=M}^{M+N-1} a_n b_n\Big| \geq \|\mathbf{a}\|_{U(k)} - \delta\ .$$

Proof. We can assume without loss that $\|\mathbf{a}\|_{U(k)} > \delta$. Let $\mathbf{I}$ and $\mathbf{c}$ be as in Proposition 6.2, but with $\delta/3$ instead of δ; we write $c_n = g(S^n y_0)$ for $n \in \mathbb{Z}$, where (Y, S, ν) is an ergodic $(k-1)$-step nilsystem, $y_0 \in Y$, and g is a smooth function on Y. We define $h = \|\!|g|\!\|_k^{-2^k+1}.\mathcal{D}_k g$ and $\mathbf{b}$ to be the sequence given by $b_n = h(S^n y_0)$, and we check that the announced properties are satisfied.

By Corollary 5.4, h is a smooth function and $\|\!|h|\!\|_{\nu,k}^* = 1$ by (14) and thus $\|\!|\mathbf{b}|\!\|_k^* = 1$. We have

$$\begin{aligned}\lim \text{averages}_{\mathbf{I}}\big(c_n b_n\big) = \lim \text{averages}_{\mathbf{I}}\big(g(S^n y_0)h(S^n y_0)\big) = \int g.h\, d\nu = \|\!|g|\!\|_k \\ = \|\mathbf{c}\|_{\mathbf{I},k} \geq \|\mathbf{a}\|_{\mathbf{I},k} - \delta/3 \geq \|\mathbf{a}\|_{U(k)} - 2\delta/3\ .\end{aligned}$$

On the other hand, by the Direct Theorem 2.13,

$$\text{limsup}\,|\text{averages}_{\mathbf{I}}((a_n - c_n)b_n)| \leq \|\mathbf{a}-\mathbf{c}\|_{\mathbf{I},k}\, \|\!|\mathbf{b}|\!\|_k^* \leq \delta/3$$

and we conclude that the lim inf of the averages on $\mathbf{I}$ of $a_n b_n$ is $\geq \|\mathbf{a}\|_{U(k)} - \delta$ and we are done. □

6.3. **Proof of Theorem 2.19.** We recall the statement for convenience.

Theorem. *For a bounded sequence $\mathbf{a} = (a_n : n \in \mathbb{Z})$, the following are equivalent.*

(i) *For every $\delta > 0$, the sequence $\mathbf{a}$ can be written as $\mathbf{a}' + \mathbf{a}''$, where $\mathbf{a}'$ is a $(k-1)$-step nilsequence, and $\|\mathbf{a}''\|_{U(k)} < \delta$.*
(ii) *For every $(k-1)$-step nilsequence $\mathbf{c} = (c_n : n \in \mathbb{Z})$, the averages of $a_n c_n$ converge.*

We recall that property (ii) means that the averages

$$\frac{1}{|I_j|}\sum_{n\in I_j} a_n c_n$$

converge for every sequence of intervals $\mathbf{I} = (I_j : n \geq 1)$ whose lengths tend to infinity. The common value of these limits is written $\lim \text{averages}\,\big(a_n c_n\big)$.

Proof. (i) $\Longrightarrow$ (ii) This implication is given by Corollary 5.11.

(ii) $\Longrightarrow$ (i)

Assume that the sequence $\mathbf{a}$ satisfies (ii). Let $\mathbf{b}$ and $\mathbf{I}$ be as in Proposition 6.2, but with $\delta/3$ instead of δ. Define $\mathbf{a}' = \mathbf{b}$ and we are left with showing that $\|\mathbf{a}-\mathbf{b}\|_{U(k)} < \delta$.

Assume that this does not hold. By Theorem 2.16, there exists a $(k-1)$-step smooth nilsequence $\mathbf{c}$ and a sequence of intervals $\mathbf{J}$ whose lengths tend to infinity with

$$\|\!|\mathbf{c}|\!\|_k^* = 1 \text{ and } \big|\lim \text{averages}_{\mathbf{J}}\big((a_n-b_n)c_n\big)\big| \geq 2\delta/3 \ .$$

Now, the sequence (b_nc_n) is a product of two $(k-1)$-step nilsequences and thus it is also a $(k-1)$-step nilsequence and its averages converge. By hypothesis, the averages of the sequence (a_nc_n) converge, and thus the averages of the sequence $(a_n-b_n)c_n$ converge. Since $\mathbf{I}$ and $\mathbf{J}$ are sequences of intervals whose lengths tend to infinity,

$$\begin{aligned}\big|\lim \text{averages}_{\mathbf{I}}\big((a_n-b_n)c_n\big)\big| &= \big|\lim \text{averages}\ \big((a_n-b_n)c_n\big)\big| \\ &= \big|\lim \text{averages}_{\mathbf{J}}\big((a_n-b_n)c_n\big)\big| \geq 2\delta/3 \ .\end{aligned}$$

On the other hand, by the Direct Theorem (Theorem 2.13)

$$\big|\lim \text{averages}_{\mathbf{I}}\big((a_n-b_n)c_n\big)\big| \leq \|\mathbf{a}-\mathbf{b}\|_{\mathbf{I},k}\, \|\!|\mathbf{c}|\!\|_k^* < 2\delta/3$$

and we have a contradiction. □

7. An application in ergodic theory

7.1. **Proof of Theorem 2.22.** We now turn to the generalization of the Wiener-Wintner Ergodic Theorem, replacing the exponential sequence $e(nt)$ by an arbitrary nilsequence. Throughout this Section, for each integer $N \geq 1$, we write I_N for the interval $[0, N-1]$ and we let $\mathbf{I}$ denote the sequence of intervals $(I_N \colon N \geq 1)$.

Let (X, μ, T) be an ergodic system, ϕ be a bounded measurable function on X, and fix an integer $k \geq 2$. We build a subset X_0 of full measure of X on which the conclusion of the Theorem holds for every $(k-1)$-step nilsequence $\mathbf{b}$.

For every integer $r \geq 1$, Corollary 3.12 of the Structure Theorem provides a $(k-1)$-step nilsystem (Z_r, ν_r, S_r), a factor map $\pi_r \colon X \to Z_r$ and a continuous function f_r on Z_r such that

$$\|\!|\phi - f_r \circ \pi_r|\!\|_k < r^{-1} \ .$$

By Corollary 3.10, there exists a subset E_r of X with $\mu(E_r) = 1$ such that for every $x \in E_r$, we have

$$\|(\phi(T^nx) - f_r\circ\pi_r(T^nx) \colon n \in \mathbb{Z})\|_{\mathbf{I},k} = \|\!|\phi - f_r\circ\pi_r|\!\|_k \leq r^{-1} \ .$$

Note that we consider the map π_r to be defined everywhere. For μ-almost every x, we have that $f_r \circ \pi_r(T^nx) = f_r(S_r^n\pi_r(x))$ for every $n \in \mathbb{Z}$. Therefore, there exists a set $E'_r \subset X$ with $\mu(E'_r) = 1$ such that

$$\|(\phi(T^nx) - f_r(S_r^n\pi_r(x)) \colon n \in \mathbb{Z})\|_{\mathbf{I},k} = \|\!|\phi - f_r\circ\pi_r|\!\|_k \leq r^{-1}$$

for every $x \in E'_r$.

Set $X_0 = \bigcap_{r=1}^\infty E'_r$. For every $x \in X_0$, the sequence $(\phi(T^nx) \colon n \in \mathbb{Z})$ satisfies the hypothesis of Corollary 5.10, completing the proof. □

7.1.1. *Proof of Corollary 2.23.* Let (X, μ, T) be an ergodic system, ϕ be a bounded measurable function on X, and let X_0 be the subset of X introduced in Theorem 2.22. Let $x \in X_0$ and p be a generalized polynomial.

For every $n \in \mathbb{Z}$, let $\{p(n)\}$ denote the fractional part of $p(n)$. Then $\{p(\cdot)\}$ is a bounded generalized polynomial. In [BL] (Theorem A, (ii)), it is shown that there exist an ergodic nilsystem (Y, ν, S), a point $y \in Y$, and a Riemann integrable function f on Y with $\{p(n)\} = f(S^ny)$ for every $n \in \mathbb{Z}$.

For every $\delta > 0$, there exists a continuous function g on Y with $\|f-g\|_{L^1(\nu)} \le \delta$. The sequence $(g(S^n y)\colon n \in \mathbb{Z})$ is a nilsequence and thus by definition of X_0, the averages on $\mathbf{I}$ of $\phi(T^n x)g(S^n y)$ converge. On the other hand, since the function $|f-g|$ is Riemann integrable and (Y,S) is uniquely ergodic, we have that

$$\begin{aligned}&\limsup|\text{averages}_{\mathbf{I}}(\phi(T^n x)(f(S^n y)-g(S^n y)))|\\&\qquad\le \|\phi\|_\infty \lim \text{averages}_{\mathbf{I}}\bigl(|f(S^n y)-g(S^n y)|\bigr) = \|\phi\|_\infty \int |f-g|\,d\nu \le \|\phi\|_\infty \delta\ .\end{aligned}$$

Therefore the averages on $\mathbf{I}$ of $\phi(T^n x)\{p(n)\} = \phi(T^n x)f(S^n y)$ form a Cauchy sequence.

We remark that for every $n \in \mathbb{Z}$, we have that $e(p(n)) = e(\{p(n)\}) = e(f(S^n y))$ and that the function $e(f(\cdot))$ is Riemann integrable on Y. The same proof gives the second claim of the corollary. □

7.2. **Examples.** Similar methods can be used to show show that some explicit sequences satisfy the hypothesis (i) of Theorem 2.19 and thus are universally good for the convergence in norm of multiple ergodic averages.

Proposition 7.1. *Let (X,T) be a uniquely ergodic system with invariant measure μ and let $k \ge 2$ be an integer. Let (Z_k, μ_k, T) be the factor defined in the Structure Theorem (Theorem 3.4) and assume that the factor map $\pi_k\colon X \to Z_{k-1}$ is continuous. Let f be a Riemann integrable function on X and let $x \in X$. Then the sequence $(f(T^n x)\colon n \in \mathbb{Z})$ satisfies hypothesis* (i) *of Theorem 2.19.*

Proof. Let $\mathbf{a}$ be the sequence $(f(T^n x)\colon n \in \mathbb{Z})$ and let $\delta > 0$. We want to show that we can write $\mathbf{a} = \mathbf{a}' + \mathbf{a}''$ where $\mathbf{a}'$ us a $(k-1)$-step nilsequence and $\|\mathbf{a}''\|_{U(k)} < \delta$.

Let (Y, S, ν), $p\colon X \to Y$, and h be the $(k-1)$-step nilsystem, the factor map, and the function on Y given by Corollary 3.12. Recall that Z_k is the inverse limit (in both the topological and measure theoretical senses) of all factors of X which are $(k-1)$-step nilsystems [HK1]. Thus Y is a factor of Z_k and the factor map $q\colon Z_k \to Y$ is continuous. Therefore the factor map $p = q \circ \pi_k$ mapping $X \to Y$ is continuous.

We define the sequences $\mathbf{a}'$ and $\mathbf{a}''$ by $a'_n = h \circ p(T^n x)$ and $a''_n = f(T^n x) - h \circ p(T^n x)$ for every $n \in \mathbb{Z}$. Then $\mathbf{a}'$ is a $(k-1)$-step nilsequence. Since the function $h \circ p$ is continuous, the function $f - h\circ p$ is Riemann integrable, and Corollary 3.11 implies that $\|\mathbf{a}''\|_{U(k)} = \| f - h \circ p \|_k < \delta$. □

We use this proposition to prove Proposition 2.20 on generalized polynomials.

Proof of Proposition 2.20. Let p be a generalized polynomial. For every $n \in \mathbb{Z}$, let $\{p(n)\}$ denote the fractional part of $p(n)$. We begin with the same argument as in the proof of Corollary 2.23.

There exists an integer $\ell \ge 1$, an ergodic ℓ-step nilsystem $(X = G/\Gamma, \mu, T)$, a point $x \in X$, and a Riemann integrable function f on X with $\{p(n)\} = f(T^n x)$ and $e(p(n)) = e(\{p(n)\}) = e(f(T^n x))$ for every $n \in \mathbb{Z}$.

The system (X, μ, T) satisfies the hypotheses of Proposition 7.1. Indeed, for $k > \ell$ we have that $Z_k = X$ and for $k < \ell$, Z_k is the quotient $G/G_k\Gamma$ of X. The result follows. □

We now prove Proposition 2.21, which states that the Thue-Morse sequence satisfies also the hypothesis of Theorem 2.19.

Proof of Proposition 2.21. Let $\mathbf{a} = (a_n : n \in \mathbb{Z})$ be the Thue-Morse sequence. We recall some of its properties (see [Q]).

There exists a uniquely ergodic system (X, T, μ), a point $x_0 \in X$, and a continuous function ϕ on X with $a_n = \phi(T^n x_0)$ for every $n \in \mathbb{Z}$. Moreover, the factor map $\pi_1 : X \to Z_1$ on the Kronecker factor Z_1 of X is continuous. Finally, the map π is two to one almost everywhere.

For every integer $k \geq 2$, the factor Z_k of X, as given by the Structure Theorem, is an extension of Z_{k-1} by a connected compact abelian group [HK1]. It follows that $Z_k = Z_1$ for every k.

Therefore the hypotheses of Proposition 7.1 are satisfied and we are done. □

7.3. **Proof of Theorem 2.24.** We now prove the generalization of the spectral theorem. Starting with an arbitrary measure preserving system (Y, S, ν), by ergodic decomposition we can assume that (Y, S, ν) is an ergodic system.

We recall the following result from [HK1] (Theorem 12.1):

Theorem. *Let $g_0, \dots, g_{k-1}$ be measurable functions on (Y, S, ν) with $\|g_i\|_\infty \leq 1$ for $i \in \{0, \dots, k-1\}$. Then*

$$\limsup_{N\to+\infty} \Big| \frac{1}{N} \sum_{n=0}^{N-1} \int \prod_{i=0}^{k-1} S^{in} g_i \, d\nu \Big| \leq c \min_{i \in \{0,\dots,k-1\}} |\!|\!| g_i |\!|\!|_{k-1}$$

where c is a constant depending only on k.

Proceeding as in [BHK] (proof of Corollary 4.5 from Theorem 4.4), we deduce:

Corollary 7.2. *Let $g_0, \dots, g_{k-1}$ be measurable functions on (Y, S, ν) with $\|g_i\|_\infty \leq 1$ for $i \in \{0, \dots, k-1\}$. Then*

$$\limsup_{N\to+\infty} \frac{1}{N} \sum_{n=0}^{N-1} \Big| \int \prod_{i=0}^{k-1} S^{in} g_i \, d\nu \Big|^2 \leq c^2 \min_{i \in \{0,\dots,k-1\}} |\!|\!| g_i |\!|\!|_k^2 \ .$$

We deduce:

Corollary 7.3. *Let $f_1, \dots, f_k$ be bounded functions on (Y, S, ν) with $\|f_i\|_\infty \leq 1$ for $i \in \{1, \dots, k\}$ and let $\mathbf{a} = (a_n : n \in \mathbb{Z})$ be a sequence with $\|\mathbf{a}\|_\infty \leq 1$. Then*

$$\limsup_{N\to+\infty} \Big\| \frac{1}{N} \sum_{n=0}^{N-1} a_n \prod_{i=1}^{k} S^{in} f_i \Big\|_{L^2(\nu)} \leq k^{1/4} c^{1/2} \min_{i \in \{1,\dots,k\}} |\!|\!| f_i |\!|\!|_{k+1} \ . \tag{18}$$

Proof. Let $\ell \in \{1, \dots, k\}$ be such that $\|f_\ell\|_{k+1} = \min_{i \in \{1,\dots,k\}} |\!|\!| f_i |\!|\!|_{k+1}$ and let Q be the $\limsup$ in the left hand side of (18).

By the van der Corput Lemma (Appendix A):

$$Q^2 \leq \limsup_{M\to+\infty} \frac{1}{M} \sum_{m=0}^{M-1} \Big| \limsup_{N\to+\infty} \frac{1}{N} \sum_{n=0}^{N-1} \overline{a_n} a_{n+m} \int \prod_{i=1}^{k} S^{in} (\overline{f_i} . S^{im} f_i) \, d\nu \Big| \ .$$

By the Cauchy-Schwarz Inequality,

$$\begin{aligned} Q^4 &\leq \limsup_{M\to+\infty} \frac{1}{M} \sum_{m=0}^{M-1} \limsup_{N\to+\infty} \frac{1}{N} \sum_{n=0}^{N-1} \Big| \int \prod_{i=1}^{k} S^{in} (\overline{f_i} . S^{im} f_i) \, d\nu \Big|^2 \\ &= \limsup_{M\to+\infty} \frac{1}{M} \sum_{m=0}^{M-1} \limsup_{N\to+\infty} \frac{1}{N} \sum_{n=0}^{N-1} \Big| \int \prod_{i=0}^{k-1} S^{in} (\overline{f_{i+1}} . S^{(i+1)m} f_{i+1}) \, d\nu \Big|^2 \ . \end{aligned}$$

Applying Corollary 7.2 to the functions $g_i = \overline{f_{i+1}}.S^{(i+1)m}f_{i+1}$, we have that

$$Q^4 \leq c^2 \limsup_{M\to+\infty} \frac{1}{M}\sum_{m=0}^{M-1} \|\overline{f_\ell}.S^{\ell m}f_\ell\|_k^2 \leq kc^2 \limsup_{M\to+\infty} \frac{1}{kM}\sum_{m=0}^{kM-1} \|\overline{f_\ell}.S^{m}f_\ell\|_k^2$$
$$\leq kc^2 \Big(\limsup_{M\to+\infty} \frac{1}{kM}\sum_{m=0}^{kM-1} \|\overline{f_\ell}.S^{m}f_\ell\|_k^{2^k}\Big)^{1/2^{k-1}}$$

by the Hölder Inequality. By (3), the last lim sup is actually a limit and is equal to $\|f_\ell\|_{k+1}^4$ and we are done. □

We now return to the proof of Theorem 2.24. We assume that $\mathbf{a} = (a_n : n \in \mathbb{Z})$ is a bounded sequence such that the averages

$$\frac{1}{N}\sum_{n=0}^{N-1} a_n b_n$$

converge as $N \to +\infty$ for every k-step nilsequence $\mathbf{b} = (b_n : n \in \mathbb{Z})$. We assume that (Y, S, ν) is an ergodic system and $f_1, \dots, f_k \in L^\infty(\nu)$. We show the convergence of the averages

$$\frac{1}{N}\sum_{n=0}^{N-1} a_n S^n f_1 \dots S^{kn} f_k$$

in $L^2(\nu)$.

Let Z_k be the k-th factor of (Y, S, ν), as given by the Structure Theorem. If for some $i \in \{1, \dots, k\}$ we have $\mathbb{E}(f_i \mid Z_k) = 0$, then $\|f_i\|_{k+1} = 0$. Then by Corollary 7.3, the above averages converge to zero in $L^2(\nu)$. We say that the *factor Z_k is characteristic for the convergence of these averages.*

Therefore, in order to prove the convergence of these averages, for arbitrary bounded functions, it suffices to prove the convergence when the functions are measurable with respect to the factor Z_k.

By the Structure Theorem, Z_k is an inverse limit of k step nilsystem. Thus by density, we can assume that the functions f_i are measurable with respect to a k-step nilsystem (Z, S) which is a factor of (Y, S, ν). By density again, we are reduced to the case that (Y, ν, S) is a k-step nilsystem and that the functions $f_1, \dots, f_k$ are continuous.

But in this case, for every $y \in Y$ the sequence

$$(f_1(S^n y).f_2(S^{2n}y).\cdots.f_k(S^{kn}y) : n \in \mathbb{Z})$$

is a k-step nilsequence and by hypothesis, the averages

$$\frac{1}{N}\sum_{n=0}^{N-1} a_n\, f_1(S^n y).f_2(S^{2n}y).\cdots.f_k(S^{kn}y)$$

converge for every $y \in Y$. □

Appendix A. The van der Corput Lemma

We state the van der Corput Lemma, as used in our set up (see [KN]):

van der Corput's Lemma. *Let* $\mathbf{a} = (a_n : n \in \mathbb{Z})$ *be a sequence with* $|a_n| \leq 1$ *for all* $n \in \mathbb{Z}$ *and let* I *be an interval in* $\mathbb{Z}$. *Then for every integer* $H \geq 1$, *we have*

$$\Big|\frac{1}{|I|}\sum_{n\in I} a_n\Big|^2 \leq \frac{4H}{|I|} + \Big|\sum_{h=-H}^{H} \frac{H-|h|}{H^2}\ \frac{1}{|I|}\sum_{n\in I} a_{n+h}\overline{a_n}\Big| \ .$$

Appendix B. Parallelepipeds in nilmanifolds

We explain the cubic structure associated to a nilmanifold. In the literature, there are (at least) two presentations of these objections, in [HK1] and in Appendix E of [GT2]. The results proved in these papers are often recalled here without proof, but we need a bit more than just those results. We use the notation of [HK1]. The group that we denote by $G^{[k]}_{k-1}$ is the same as the group HP^k of [GT2].

The k's in index and exponent that occur everywhere are cumbersome but necessary as we use an induction at some point.

B.1. **Algebraic preliminaries.**

We begin with some algebraic constructions involving "cubes." Let G be a group and $k \geq 1$ be an integer.

B.1.1. *Two constructions of the "side group".*

We use the notation of Section 3.2. We write $\mathbf{0} = (0, 0, \dots, 0) \in \{0, 1\}^k$ and $\mathbf{1} = (1, 1, \dots, 1) \in \{0, 1\}^k$.

As before, if X is a set, $X^{[k]} = X^{2^k}$ and points of $X^{[k]}$ are written as $\underline{x} = (x_\epsilon : \epsilon \in \{0, 1\}^k)$. For $x \in X$, $x^{[k]} \in X^{[k]}$ is the element $(x, x, \dots, x)$, with x repeated 2^k times. If $f : X \to Y$ is a map, $f^{[k]} : X^{[k]} \to Y^{[k]}$ denotes the diagonal map: $(f(\underline{x}))_\epsilon = f(x_\epsilon)$ for all $\epsilon \in \{0, 1\}^k$.

For $g \in G$ and $1 \leq i \leq k$, $g_i^{[k]} = ((g_i^{[k]})_\epsilon : \epsilon \in \{0, 1\}^k)$ is given by:

$$(g_i^{[k]})_\epsilon = \begin{cases} g & \text{if } \epsilon_i = 1 \\ 1 & \text{if } \epsilon_i = 0 \ . \end{cases}$$

(Note that we mean $\epsilon = (\epsilon_1, \dots, \epsilon_k)$.) $G^{[k]}_{k-1}$ is the subgroup of $G^{[k]}$ spanned by

$$\{g^{[k]} : g \in G\} \cup \{g_i^{[k]} : 1 \leq i \leq k,\ g \in G\} \ .$$

The same group was also introduced in [GT2], but with a different definition and notation. We recall their presentation, but in our notation, substituting "upper faces" for "lower faces" for coherence. We start with some notation.

It is convenient to view $\{0, 1\}^k$ as the set of vertices of the unit Euclidean cube. If J is a subset of $\{1, \dots, k\}$ and $\eta \in \{0, 1\}^J$, the set

$$\alpha = \{\epsilon \in \{0, 1\}^k : \epsilon_i = \eta_i \text{ for all } i \in J\}$$

is called a *face* of $\{0, 1\}^k$. The dimension of α is $\dim(\alpha) = k - |J|$. If all coordinates of η are equal to 1, then this face is called an *upper face*. In particular, $\alpha_0 = \{0, 1\}^k$ is the unique upper face of dimension k, corresponding to $J = \emptyset$; $\{\mathbf{1}\}$ is the unique upper face of dimension zero, corresponding to $J = \{1, \dots, k\}$. The k upper faces of dimension $k - 1$ are $\alpha_i = \{\epsilon \in \{0, 1\}^k : \epsilon_i = 1\}$ for $1 \leq i \leq k$. Let $\alpha_0, \alpha_1, \dots, \alpha_{2^k}$ be an enumeration of all of the upper faces such that $\alpha_0, \dots, \alpha_k$ are as above and $\dim(\alpha_i)$ is a decreasing sequence; in particular, $\alpha_{2^k} = \{\mathbf{1}\}$.

If α is a face and $g \in G$, we write $g_\alpha^{[k]} = ((g_\alpha^{[k]})_\epsilon \colon \epsilon \in \{0,1\}^k)$ for the element of $G^{[k]}$ given by:

$$\big(g_\alpha^{[k]}\big)_\epsilon = \begin{cases} g & \text{if } \epsilon \in \alpha \ ; \\ 1 & \text{otherwise} \ . \end{cases}$$

In particular, the elements $g_i^{[k]}$ defined above can be written as $g_{\alpha_i}^{[k]}$.

In [GT2], $\mathrm{HP}^k(G)$ is defined to be the set of elements $\underline{g} \in G^{[k]}$ that can be written as

(19) $\underline{g} = (g_1)_{\alpha_1}^{[k]}(g_2)_{\alpha_2}^{[k]} \dots (g_{2^k})_{\alpha_{2^k}}^{[k]}$ where $g_i \in G_{k-\dim(\alpha_i)}$ for every $i \in \{1, \dots, k\}$.

Here $G_0 = G_1 = G$; in all other places in the paper, we use G_0 to denote a different object (the connected component of the identity of G).

Let us explain briefly why $G_{k-1}^{[k]}$ and $\mathrm{HP}^k(G)$ are actually equal. By a direct computation, Green and Tao show that $\mathrm{HP}^k(G)$ is a subgroup of $G^{[k]}$; since it contains the generators of $G_{k-1}^{[k]}$, it contains this group. On the other hand, it is shown in [HK1] (section 5) that for every side α of dimension d and every $g \in G_{k-\dim(\alpha)}$, $g_\alpha^{[k]}$ belongs to $G_{k-1}^{[k]}$ (and more precisely to $(G_{k-1}^{[k]})_{k-\dim(\alpha)}$) and thus $\mathrm{HP}^k(G) \subset G_{k-1}^{[k]}$. We have equality.

In the sequel we only use the notation $G_{k-1}^{[k]}$. Depending on the property to be proven, the first or second presentation is more convenient.

B.1.2. *Algebraic properties.* We have:

(i) Let Γ be a subgroup of G. If all coordinates of $\underline{g}$ belong to Γ except possibly $g_{\mathbf{0}}$, then $g_{\mathbf{0}} \in \Gamma G_k$.

(ii) In particular, if all coordinates of $\underline{g} \in G_{k-1}^{[k]}$ are equal to 1 except possibly $g_{\mathbf{0}}$, then $g_{\mathbf{0}} \in G_k$.

The second statement is proved (in a perhaps concealed place) in [HK1] via induction on k, and the first one is not stated explicitly but follows with a similar proof. Both statements follow easily from the second definition of $G_{k-1}^{[k]}$ and the symmetry of this set, allowing us to substitute the coordinate $g_{\mathbf{1}}$ for $g_{\mathbf{0}}$.

We need two more groups for our proofs. In this appendix, we write

$$H_k = \{\underline{g} \in G_{k-1}^{[k]} \colon g_{\mathbf{0}} = 1\} \text{ and } G_k^{[k]} = \{g^{[k]} \colon g \in G\} \ .$$

(The first group is not defined in the papers.) Then H_k is clearly a normal subgroup of $G_{k-1}^{[k]}$ and $G_{k-1}^{[k]} = H_k.G_k^{[k]}$. Moreover, H_k is the group spanned by the elements $g_i^{[k]}$ for $1 \le i \le k$ and $g \in G$; in the second presentation of $G_{k-1}^{[k]}$, it consists of elements that can be written as in (19) with $g_1 = 1$.

We have

(iii) $(H_k)_2 = H_k \cap (G_2)^{[k]}$.

(iv) $(G_{k-1}^{[k]})_2 = G_{k-1}^{[k]} \cap (G_2)^{[k]}$.

Proof. We prove (iii). The inclusion $(H_k)_2 \subset H_k \cap (G_2)^{[k]}$ is obvious.

Let α be a face of dimension $d < k-1$ containing $\mathbf{1}$. Let $g \in G$ and $h \in G_{k-d-1}$. We can chose a face β of dimension $k-1$ and a face γ of dimension $d+1$ such that $\alpha = \beta \cap \gamma$. We have

$$g_\alpha^{[k]} \in H_k \ ; \ h_\gamma^{[k]} \in H_k \text{ and } [g;h]_\alpha^{[k]} = \big[g_\beta^{[k]}; h_\gamma^{[k]}\big] \ .$$

Thus $[g;h]^{[k]}_\alpha \in (H_k)_2$. Therefore, for any $q \in G_{k-d}$, we have that $q^{[k]}_\alpha \in (H_k)_2$.

Using this remark, we can show the inclusion $H_k \cap (G_2)^{[k]} \subset (H_k)_2$. Let $\underline{g}$ be in the first of these groups. We write $\underline{g}$ as in (19) with $g_1 = 1$. By the remark, all terms of the form $(g_j)^{[k]}_{\alpha_j}$ with $\dim(\alpha_j) < k-1$ in the product belong to $(H_k)_2$ and we are reduced to show that the product of the k remaining terms also belongs to this group. We remark that all coordinates of this product belong to G_2.

Let $g^{[k]}_\alpha$ be one of these terms. Then α is an upper face of dimension $k-1$ and it is immediate that there exists $\eta \in \{0,1\}^k$ such that η belongs to α and does not belong to any other upper face of dimension $k-1$. Therefore, g is the coordinate η of the product and $g \in G_2$. It follows that $g^{[k]}_\alpha$ belongs to $(H_k)_2$ and we are done.

We now deduce (iv). Again, the inclusion $(G^{[k]}_{k-1})_2 \subset G^{[k]}_{k-1} \cap (G_2)^{[k]}$ is obvious. Let $\underline{g} \in G^{[k]}_{k-1} \cap (G_2)^{[k]}$. We write $\underline{g} = h^{[k]}\underline{q}$ where $h \in G$ and $\underline{q} \in H_k$. We have that $g_0 = h$ and so $h \in G_2$. Thus $h^{[k]} \in (G_2)^{[k]}$. Moreover, $\underline{q} \in H_k \cap (G_2)^{[k]}$ and by the second part of the Lemma, $\underline{q} \in (H_k)_2 \subset (G_2)^{[k]}_{k-1}$. □

B.2. **Topological properties.** Henceforth G is a r-step nilpotent Lie group, Γ is a discrete cocompact subgroup, and $X = G/\Gamma$. In applications r will be equal to $k-1$ but the general case is used in an induction below.

In [HK1] and [GT2], it is shown that

(v) $G^{[k]}_{k-1}$ is a closed subgroup of $G^{[k]}$ and hence is an r-step nilpotent Lie group.

(vi) The group $\Lambda_k := \Gamma^{[k]} \cap G^{[k]}_{k-1}$ is a cocompact subgroup of $G^{[k]}_{k-1}$.

We do not reproduce the proof here. We define:

$$X_k = G^{[k]}_{k-1}/(\Gamma^{[k]} \cap G^{[k]}_{k-1}) \ .$$

For the moment we write ν_k for the Haar measure of X_k.

The image of ν_k under the projection $\underline{x} \mapsto x_{\mathbf{0}}$ is equal to the Haar measure μ of X. We have that:

(vii) The group $\Theta_k := H_k \cap \Gamma^{[k]}$ is cocompact in H_k.

Proof. Every $\underline{g} \in H_k$ belongs to $G^{[k]}_{k-1}$ and thus is at a bounded distance from some $\underline{\gamma} \in \Lambda_k$. Since $g_{\mathbf{0}} = 1$, $\gamma_{\mathbf{0}}$ is at a bounded distance from 1. Since Γ is discrete, $\gamma_{\mathbf{0}}$ belongs to a finite subset F of Γ.

We have that $\underline{g}$ is at a bounded distance from $((\gamma_{\mathbf{0}})^{[k]})^{-1}\underline{\gamma}$, which belongs to $G^{[k]}_{k-1} \cap H_k = \Theta_k$. □

We define $W_k = H_k/\Theta_k$. Then W_k is a $(k-1)$-step nilmanifold, naturally included in X_k as a closed subset.

For every $g \in G$ we have that $g^{[k]}$ belongs to $G^{[k]}_{k-1}$. We deduce that for every $x \in X$, we have that $x^{[k]} := (x, x, \dots, x)$ belongs to X_k.

For every $x \in X$, we write

$$W_{k,x} = \{\underline{x} \in X_k \colon x_{\mathbf{0}} = x\} \ .$$

We show:

(viii) Let $x \in X$ and g be a lift of x in G. Then $W_{k,x} = g^{[k]}.W_k$.

Proof. Let $\underline{x} \in W_{k,x}$ and $\underline{h}$ be a lift of $\underline{x}$ in $G^{[k]}_{k-1}$. Since $x_{\mathbf{0}} = x$, we have that $h_{\mathbf{0}} = g\gamma$ for some $\gamma \in \Gamma$. Let $\underline{q} = (g^{[k]})^{-1}\underline{h}(\gamma^{[k]})^{-1}$. Then $\underline{q} \in H_k$ and its image $\underline{y}$ in H_k satisfies $g^{[k]}\underline{y} = \underline{x}$. We thus have that $W_{k,x} \subset g^{[k]}.W_k$ and the opposite inclusion is obvious. □

B.3. **Dynamical properties.** Henceforth, we assume that X is endowed with the translation T by some $\tau \in G$ and that (X, T, μ) is ergodic. Recall that the same nilmanifold can be represented as a quotient in different ways. As usual we assume that G is spanned by the connected component G_0 of the identity and τ. We claim that:

(ix) $(G^{[k]}_{k-1})_0 = (G_0)^{[k]}_{k-1}$.
(x) $G^{[k]}_{k-1}$ is spanned by $(G^{[k]}_{k-1})_0$, $\tau^{[k]}$, and the elements $\tau^{[k]}_i$, $1 \leq i \leq k$.
(xi) H_k is spanned by $(H_k)_0$ and the elements $\tau^{[k]}_i$, $1 \leq i \leq k$.

Proof. By hypothesis and the first definition of $G^{[k]}_{k-1}$, this group is spanned by elements of the form $g^{[k]}$ for $g \in G_0$, $g^{[k]}_i$ for $g \in G_0$ and $1 \leq i \leq k$, $\tau^{[k]}_i$ for $1 \leq i \leq k$ and $\tau^{[k]}$. This proves (x).

The commutator of two elements of the above type belongs to $(G_2)^{[k]}_{k-1} \subset (G_0)^{[k]}_{k-1}$, because it follows from our assumption that $G_2 \subset G_0$. Then every element $\underline{g}$ of $G^{[k]}_{k-1}$ can be written as $\underline{g} = \underline{h}(\tau^{[k]})^n(\tau^{[k]}_1)^{m_1} \dots (\tau^{[k]}_k)^{m_k}$ with $\underline{h} \in (G_0)^{[k]}_{k-1}$.

If $\underline{g} \in (G^{[k]}_{k-1})_0$, then by looking at the coordinate $\mathbf{0}$ of $\underline{g}$ we have that $h_0\tau^n = g_0$ belongs to G_0. Thus $\tau^n \in G_0$.

Let $i \in \{1, \dots, k\}$. As in the proof of (iii), there exists $\eta \in \{0,1\}^k$ such that $\tau^{[k]}_i = \tau$ and $\tau^{[k]}_j = 1$ for $j \neq i$. We have that $g_\eta = h_\eta\tau^{m_i}_i$ and thus $\tau^{m_i} \in G_0$. Thus $(\tau^{[k]}_i)^{m_i} \in (G_0)^{[k]}_{k-1}$. This achieves the proof of (ix).

Now assume that $\underline{g} \in (H_k)_0$. Then it belongs to $(G^{[k]}_{k-1})_0$ and we write it as above, $\underline{g} = \underline{h}(\tau^{[k]}_1)^{m_1} \dots (\tau^{[k]}_k)^{m_k}$ with $\underline{h} \in (G_0)^{[k]}_{k-1}$. We have that $h_{\mathbf{0}} = g_{\mathbf{0}} = 1$ and so $\underline{h} \in H_k \cap (G_0)^{[k]}_{k-1}$ and this element belongs to $(H_k)_0$. This proves (xi). □

(xii) X_k is ergodic under the action of $T^{[k]}$ and $T^{[k]}_i$, $1 \leq i \leq k$.
(xiii) W_k is ergodic under the transformations $T^{[k]}_i$, $1 \leq i \leq k$.

Proof. Let Z be the compact abelian group $G/\Gamma G_2$ and σ be the image of τ in Z. Since T is ergodic, the translation by σ on Z is ergodic.

By (iv) and (any) definition of $G^{[k]}_{k-1}$, the quotient $G^{[k]}_{k-1}/(G^{[k]}_{k-1})_2\Lambda_k$ can be identified with the subgroup $Z^{[k]}_{k-1}$ of $Z^{[k]}$. This group consists of the points $\underline{z}$ of $Z^{[k]}$ which can be written as

$$\underline{z} = \big(u\prod_{i=1}^k v_i^{\epsilon_i} : \epsilon \in \{0,1\}^k\big)$$

for some $u, v_1, \dots, v_k \in Z$. The transformations induced on this group by the transformations $T^{[k]}$ and $T^{[k]}_i$, $1 \leq i \leq k$, are the translations by $\sigma^{[k]}$ and $\sigma^{[k]}_i$. In the above parametrization of $Z^{[k]}_{k-1}$, these transformations correspond to the map $u \mapsto \sigma u$ and to the maps $v_i \mapsto \sigma v_i$, respectively.

Since the translation by σ on Z is ergodic, it follows easily that $Z^{[k]}_{k-1}$ is ergodic under the translations by $\sigma^{[k]}$ and $\sigma^{[k]}_i$. By (ix) and Theorem 3.6, X_k is ergodic under the action of $T^{[k]}$ and $T^{[k]}_i$, $1 \leq i \leq k$.

The second statement is proved in the same way. □

We show:

(xiv) The Haar measure ν_k of X_k is equal to the measure $\mu^{[k]}$ defined in [HK1] and described in Section 3.2.

This result is proved in [HK1], but the context is so different from the present one that we prefer to give a complete proof here.

Proof. We use induction on k. By definition, $G^{[2]}_1 = G \times G$ and so $X_1 = X \times X$ and $\nu_1 = \mu \times \mu$, which is equal to the measure μ_1 of [HK1].

Assume that the announced property holds up to $k-1$ for some $k > 1$. In order to show the property for k, it suffices to show that when f_ϵ, $\epsilon \in \{0,1\}^k$, are 2^k continuous functions on X, we have that the function F defined on $X^{[k]}$ by

$$F(\underline{x}) = \prod_{\epsilon \in \{0,1\}^k} f_\epsilon(x_\epsilon)$$

has the same integral under the measures $\mu^{[k]}$ and ν_k.

For every $x \in X$, the point $x^{[k]} = (x, x, \dots, x)$ belongs to X_k. Since $(X_k, T^{[k]}, T^{[k]}_1, \dots, T^{[k]}_k)$ is uniquely ergodic with invariant measure ν_k, we have that

$$\int F(\underline{x})\, d\nu_k(\underline{x})$$
$$= \lim_{L\to+\infty} \frac{1}{L} \sum_{\ell=0}^{L-1} \Big(\lim_{M\to+\infty} \frac{1}{M^{k-1}} \sum_{m_1,\dots m_{k-1}=0}^{M-1} \Big(\lim_{N\to+\infty} \frac{1}{N} \sum_{n=0}^{N-1} \prod_{\epsilon\in\{0,1\}^k} f_\epsilon(T^{n+\epsilon\cdot m+\epsilon_k\ell} x) \Big)\Big)$$

where $m = (m_1, \dots, m_{k-1})$ and $\epsilon\cdot m = \epsilon_1 m_1 + \ldots + \epsilon_{k-1} m_{k-1}$. By unique ergodicity of (X, T, μ), this is equal to

$$\lim_{L\to+\infty} \frac{1}{L} \sum_{\ell=0}^{L-1} \Big(\lim_{M\to+\infty} \frac{1}{M^{k-1}} \sum_{m_1,\dots m_{k-1}=0}^{M-1} \int \prod_{\epsilon\in\{0,1\}^k} f_\epsilon(T^{\epsilon\cdot m+\epsilon_k\ell} x)\, d\mu(x) \Big) \ .$$

We write each $\epsilon \in \{0,1\}^k$ in the form $\eta 0$ or $\eta 1$ with $\eta \in \{0,1\}^{k-1}$, and this expression can be rewritten as

$$\lim_{L\to+\infty} \frac{1}{L} \sum_{\ell=0}^{L-1} \Big(\lim_{M\to+\infty} \frac{1}{M^{k-1}} \sum_{m_1,\dots,m_{k-1}=0}^{M-1} \int \prod_{\eta\in\{0,1\}^{k-1}} (f_{\eta 0}.T^\ell f_{\eta 1})(T^{\eta\cdot m} x)\, d\mu(x) \Big) \ .$$

By unique ergodicity of X_{k-1} under the transformations $T^{[k-1]}$ and $T^{[k-1]}_i$, $1 \leq i \leq k-1$, and proceeding as above, we have that this expression is equal to

$$\lim_{L\to+\infty} \frac{1}{L} \sum_{\ell=0}^{M-1} \int \prod_{\eta\in\{0,1\}^{k-1}} (f_{\eta 0}.T^\ell f_{\eta 1})(x_\eta)\, d\nu_{k-1}(\underline{x}) \ .$$

By the induction hypothesis, the integral remains unchanged when the measure $\mu^{[k-1]}$ is substituted for ν_{k-1}. We rewrite this expression as

$$\lim_{L\to+\infty}\frac{1}{L}\sum_{\ell=0}^{L-1}\int F_0\,.\,F_1\circ(T^{[k-1]})^\ell\,d\mu^{[k-1]} \tag{20}$$

where

$$F_0(\underline{x})=\prod_{\eta\in\{0,1\}^{k-1}} f_{\eta 0}(x_\eta)\text{ and }F_1(\underline{x})=\prod_{\eta\in\{0,1\}^{k-1}} f_{\eta 1}(x_\eta)\ .$$

Let $\mathcal{I}$ denotes the $T^{[k-1]}$-invariant σ-algebra of the measure $\mu^{[k-1]}$. The limit (20) is equal to

$$\int \mathbb{E}(F_0\mid\mathcal{I})\,\mathbb{E}(F_1\mid\mathcal{I})\,d\mu^{[k-1]}\ .$$

By the inductive definition of the measure $\mu^{[k]}$ in [HK1] (section 3), this is equal to

$$\int F_0(x_{\eta 0}\colon \eta\in\{0,1\}^{k-1})\,F_1(x_{\eta 1}\colon \eta\in\{0,1\}^{k-1})\,d\mu^{[k]}(\underline{x})$$

and the function in the integral is just the function F. □

Recall that the measure $\mu^{[k]}$ satisfies the inequality (2) of Section 3.2. This can probably be proved directly for the measure ν_k but does not seem obvious.

B.4. **The fibers.** Recall that for every $x\in X$, $W_{k,x}=\{\underline{x}\in X_k\colon x_{\mathbf{0}}=x\}$.

(xv) For every $x\in X$, $W_{k,x}$ is uniquely ergodic under the transformations $T_i^{[k]}$, $1\leq i\leq k$.

We write ρ_x for the invariant measure of $W_{k,x}$.

(xvi) For every $x\in X$ and $h\in G$, $\rho_{h.x}$ is the image of ρ_x under the translation by $h^{[k]}$.

Proof. Let g be a lift of x in G and $\tilde\tau=g\tau g^{-1}$.

For $1\leq i\leq k$, we have that $\tilde\tau_i^{[k]}=g^{[k]}\tau_i^{[k]}(g^{[k]})^{-1}$ and all these elements commute and belong to H_k. For $1\leq i\leq k$, let $\tilde T_i^{[k]}$ be the translation by $\tilde\tau_i^{[k]}$.

We first show that the nilsystem $(W_k,\tilde T_1^{[k]},\dots,\tilde T_k^{[k]})$ is uniquely ergodic. For each i, $\tilde\tau_i^{[k]}(\tau_i^{[k]})^{-1}$ belongs to $H_k\cap(G_2)^{[k]}$ and thus to $(H_k)_2$ by (iii). Therefore, $\tilde\tau_i^{[k]}$ and $\tau_i^{[k]}$ have the same projection on the compact abelian group $H_k/(H_k)_2$. By (xiii), the action induced by $\tau_i^{[k]}$, $1\leq i\leq k$ on this group is ergodic. The criteria given by Theorem 3.6 and property (xi) give the announced unique ergodicity.

By (viii), we have that $g^{[k]}.W_k=W_{k,x}$. The map $\underline{y}\mapsto g^{[k]}.\underline{y}$ mapping $(W_k,\tilde T_1^{[k]},\dots,\tilde T_k^{[k]})$ to $(W_{k,x},T_1^{[k]},\dots,T_k^{[k]})$ is an isomorphism of topological systems and thus the second of these system is uniquely ergodic. This proves (xv).

We write ρ for the Haar measure of the nilmanifold $W_k=H_k/\Theta_k$. Then ρ is the invariant measure of W_k and the above proof shows that for every $g\in G$, the invariant measure of $W_{k,x}$ is the image of ρ under translation by $G^{[k]}$. This immediately implies (xvi). □

In fact, $W_{k,x}$ can be given the structure of a nilmanifold, quotient of the group H_k by the discrete cocompact group $g^{[k]}\Theta(g^{[k]})^{-1}$, and the transformations $T_i^{[k]}$ are translations on this nilmanifold.

B.5. **The case that G is a $(k-1)$-step nilpotent.** Henceforth we assume that G is a $(k-1)$-step nilpotent group.

We show:

(xvii) Let X_{k*} be the image of $\underline{x} \mapsto \underline{x}_*$ of X_k under the projection $\underline{x} \mapsto \underline{x}_*$ mapping $X^{[k]}$ to X^{2^k-1}. There exists a smooth map $\Phi\colon X_{k*} \to X_k$ such that

$$X_k = \left\{(\Phi(\underline{x}_*), \underline{x}_*)\colon \underline{x} \in X_{k*}\right\} \ . \tag{21}$$

Different proofs are given for the existence and continuity of Φ in [HK1] and [GT2]. The smoothness of Φ can be easily deduced from these proofs, but this property is not stated in these papers. For completeness, we give a short complete proof.

Proof. First we remark that the projection $X_k \to X_{k*}$ is one to one. Indeed, let $\underline{x}$ and $\underline{y}$ be two points of X_k with the same projections. We lift them to two elements $\underline{g}$ and $\underline{h}$ of $G^{[k]}$. All the coordinates of $\underline{hg}^{-1}$ belong to Γ except the first one, and by (i) this coordinate also belongs to $\Gamma G_k = \Gamma$. Thus $\underline{x} = \underline{y}$.

Therefore the projection $X_k \to X_{k*}$ is a homeomorphism. By composing the reciprocal of this map with the projection $\underline{x} \mapsto x_{\mathbf{0}}$, we obtain a continuous map $\Phi\colon X_{k*} \to X$ satisfying (21). We are left with showing that it is smooth.

Let G_* be the image of $G^{[k]}$ in G^{2^k-1} under the map $\underline{g} \mapsto \underline{g}_*$. By (ii), the projection $G^{[k]} \to G_*$ is one to one.

We check that G_* is a closed subgroup of G^{2^k-1}. Let $(\underline{g}_{*n})$ be a sequence in G^{2^k-1} converging to some $\underline{g}_*$. For each n, there exists $g_{\mathbf{0},n} \in G$ with $\underline{g}_n = (g_{\mathbf{0},n}, \underline{g}_{*n}) \in G^{[k]}_{k-1}$ and there exists $\underline{\gamma}_n \in \Gamma^{[k]} \cap G^{[k]}_{k-1}$ at a bounded distance from $\underline{g}_n$. All the coordinates of $\underline{\gamma}_n$, except $\gamma_{\mathbf{0}}$, are for all n at a bounded distance from the unit. By passing to subsequences, we can assume that they do not depend on n. By (i), $\underline{\gamma}_n$ does not depend on n. Therefore, $\underline{g}_n$ remains at a bounded distance from the unit and taking a subsequence we can assume that it converges to some $\underline{g}$, which belongs to $G^{[k]}_{k-1}$ by (v). Then the projection of $\underline{g}$ on G_* is equal to $\underline{g}_*$. Thus $\underline{g}$ belongs to G_*.

Now, the projection $G^{[k]}_{k-1} \to G_*$ is a smooth bijective homomorphism between Lie groups. Therefore it is a diffeomorphism. Since the projection $G^{[k]}_{k-1} \to X_k$ has discrete kernel, it follows that the projection $X_k \to X_{k*}$ is a diffeomorphism and thus that Φ is smooth. $\square$

We deduce:

(xviii) $\|\!|\cdot|\!\|_k$ is a norm on $\mathcal{C}(X)$.

Proof. It suffices to show that if $f \in \mathcal{C}(X)$ satisfies $\|\!|f|\!\|_k = 0$, then $f = 0$. By Proposition 4.3, if f_ϵ, $\epsilon \in \{0,1\}^k_*$, are 2^k-1 continuous functions on X, then

$$\int f(x_{\mathbf{0}}) \prod_{\epsilon \in \{0,1\}^k - *} f_\epsilon(x_\epsilon)\, d\mu^{[k]}(\underline{x}) = 0 \ .$$

By density, $\int f(x_0) F(\underline{x}_*)\, d\mu(\underline{x}) = 0$ for every continuous function F on X_{k*}. Taking $F = \bar{f} \circ \Phi$ where Φ is as in statement ii of Theorem 5.1, property (21) of this function gives

$$0 = \int f(x_0)\bar{f}(\Phi(x_*))\,d\mu^{[k]}(\underline{x}) = \int |f(x_0)|^2\,d\mu^{[k]}(\underline{x}) = \int |f(x)|^2\,d\mu(x)$$

because the projection of $\mu^{[k]}$ on X is μ. □

References

[A] J. Auslander. *Minimal flows and their extensions.* North Holland Publishing Co, Amsterdam, 1988.

[AGH] L. Auslander, L. Green and F. Hahn. Flows on homogeneous spaces. *Ann. Math. Studies* **53**, Princeton University Press, 1963.

[BFKO] J. Bourgain, H. Furstenberg, Y. Katznelson, D. Ornstein. Appendix on return-time sequences. *Inst. Hautes Études Sci. Publ. Math.* **69** (1989), 42-45.

[BFW] V. Bergelson, H. Furstenberg and B. Weiss. Piecewise-Bohr sets of integers and combinatorial number theory. Algorithms Combin. **26**, Springer, Berlin (2006), 13-37.

[BHK] V. Bergelson, B. Host and B. Kra, with an Appendix by I. Ruzsa. Multiple recurrence and nilsequences. *Inventiones Math.* **160** (2005), 261-303.

[BL] A. Leibman and V. Bergelson. Distribution of values of bounded generalized polynomials. *Acta Math.* **198** (2007), 155-230.

[E] R. Ellis. *Lectures on topological dynamics.* W. A. Benjamin Inc., New York, 1969.

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

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

[GT1] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. To appear, *Annals of Math.* Available at: http://arxiv.org/abs/math/0404188

[GT2] B. Green and T. Tao. Linear equations in the primes. To appear, *Annals of Math.* Available at: http://arxiv.org/abs/math/0606088

[GT3] B. Green and T. Tao. Quadratic uniformity of the Möbius function. To appear, *Annales de l'Institut Fourier.* Available at: http://arxiv.org/abs/math/0606087

[GT4] B. Green and T. Tao. An inverse theorem for the Gowers U^3-norm, with applications. To appear, *Proc. Edinburgh Math. Soc.* Available at: http://arxiv.org/abs/math/0503014

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

[HK2] B. Host and B. Kra. Analysis of two step nilsequences. Submitted. Available at: http://arxiv.org/abs/0709.3241

[KN] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences.* John Wiley and Sons, New York, 1974.

[Lei] A. Leibman. Pointwise convergence of ergodic averages for polynomial sequences of rotations of a nilmanifold. *Erg. Th. & Dynam. Sys.* **25** (2005), 201-213.

[Les1] E. Lesigne. Sur une nil-variété, les parties minimales associées à une translation sont uniquement ergodiques. *Erg. Th. & Dynam. Sys.* **11** (1991), 379-391.

[Les2] E. Lesigne. Spectre quasi-discret et théorème ergodique de Wiener-Wintner pour les polynômes. *Erg. Th. & Dynam. Sys.* **13** (1993), 767-784.

[Q] M. Queffelec. *Substitution Dynamical Systems – Spectral Analysis.* Lecture Notes in Math. **1294** Springer-Verlag, New York (1987).

[WW] N. Wiener and A. Wintner. Harmonic analysis and ergodic theory. *Amer. J. Math.* **63** (1941), 415-426.

Université Paris-Est, Laboratoire d'analyse et de mathématiques appliquées, UMR CNRS 8050, 5 bd Descartes, 77454 Marne la Vallée Cedex 2, France

E-mail address: `bernard.host@univ-mlv.fr`

Department of Mathematics, Northwestern University, 2033 Sheridan Road, Evanston, IL 60208-2730, USA

E-mail address: `kra@math.northwestern.edu`