# MINIMAL CHARACTERISTIC FACTORS FOR LINEAR SYSTEMS

J. WOLF

Abstract. In this expository note we outline the analogies between two recent preprints by Leibman [Lei07] and Gowers and the author [GW07]. Both papers independently describe two manifestations of the same phenomenon, the former in the context of ergodic theory and the latter in arithmetic combinatorics. In their respective settings, they address the question after the degree of the minimal characteristic factor of a multiple ergodic average along a system of linear forms, or the minimal degree of uniformity needed to accurately count solutions to the corresponding system of linear equations. The present article is aimed at readers with a combinatorics background and limited prior exposure to ergodic theory.

## 1. Introduction

In [GW07] we investigated the following question: for which types of systems of linear equations can we guarantee that any subset of $\mathbb{F}_p^n$ which is uniform of degree $k$ contains the "expected" number of solutions, that is, the number of solutions one would expect in a random subset of the same density. By *uniform of degree* $k$ we mean that the balanced function of the set is small in the so-called $U^{k+1}$-norm, which originated in the work of Gowers on Szemerédi's Theorem for long arithmetic progressions [Gow01] and will be defined at the start of Section 3.

To make this question more precise, we developed a new notion of complexity of a linear system which we called the *true complexity*. For example, we defined a system of linear forms $\mathcal{L} = (L_1, \ldots, L_m)$ on $(\mathbb{F}_p^n)^d$ to have *true complexity* 1 if and only if it contains the "correct" number of solutions in any uniform set. More generally, we say a system has *true complexity* $k$ if $k$ is the least integer such that the average over the linear forms is governed by the $U^{k+1}$-norm.

We then proceeded to show, under one additional assumption, that linear systems of true complexity 1 are precisely those for which the squares of the linear forms defining the system are linearly independent. It is straightforward to show that square-independence is a necessary condition for true complexity 1 by adapting a well-known construction used to

*Date*: July 2007.

show that there are uniform sets which contain too many 4-term progressions. The precise qualitative version of the fact that square-independence is also a sufficient condition for true complexity 1 (again, under one additional assumption) is best stated as follows:

**Theorem 1.1.** *For every $\epsilon > 0$ there exists $c > 0$ with the following property. Let $f : \mathbb{F}_p^n \to [-1, 1]$ satisfy $\|f\|_{U^2} \leq c$. Let $\mathcal{L} = (L_1, \ldots, L_m)$ be a square-independent system of linear forms in $d$ variables of Cauchy-Schwarz complexity at most 2. Then*

$$\left| \mathbb{E}_{x_1,\ldots,x_d \in \mathbb{F}_p^n} \prod_{i=1}^{m} f(L_i(x_1, \ldots, x_d) \right| \leq \epsilon.$$

*In other words, $\mathcal{L}$ has true complexity 1.*

The reader will of course have noticed we have not yet stated what we mean by the Cauchy-Schwarz complexity of a linear system. We shall not do so here, but refer the interested reader to [GW07] or the original paper [GT06a]. All that matters to us here is that Cauchy-Schwarz complexity is precisely the condition that enables us to prove the following theorem via a simple Cauchy-Schwarz argument:

**Theorem 1.2.** *Let $f_1, \ldots, f_m$ be functions from $\mathbb{F}_p^n$ to $[-1, 1]$, and let $\mathcal{L}$ be a linear system of Cauchy-Schwarz complexity $k$ consisting of $m$ forms in $d$ variables. Then*

$$\left| \mathbb{E}_{x_1,\ldots,x_d \in \mathbb{Z}_N} \prod_{i=1}^{m} f_i(L_i(x_1, ..., x_d)) \right| \leq \min_i \|f_i\|_{U^{k+1}}.$$

The additional hypothesis of Cauchy-Schwarz complexity 2 in Theorem 1.1 is a technical yet important condition. It stems from the fact that when considering an average such as the one in Theorem 1.1, it is convenient to decompose the function $f$ into a quadratically structured part and a part that is small in $U^3$, and then Proposition 1.2 tells us that for systems of Cauchy-Schwarz complexity 2, only the contribution from the structured part needs to be considered. Unfortunately, we do not currently have such a decomposition for higher-order $U^k$-norms, hence the restriction to systems of Cauchy-Schwarz complexity 2.

**Example 1.3.** *Linear systems that were previously thought to require quadratic Fourier analysis but that have been shown to be governed by the $U^2$-norm by Theorem 1.1 include the systems $\mathcal{L}_1 = (x, n, m, x + n + m, x + 2n - m, x + 2m - n)$, the translation-invariant $\mathcal{L}_2 = (x, x + n, x + m, x + n + m, x + n - m, x + m - n)$ and $\mathcal{L}_3 = (x, x + n, x + m, x + k, x + n + m, x + m + k, x + n + k)$, which represents a combinatorial cube of dimension 3 with one vertex missing.*

From Theorem 1.1 we deduced the following corollary concerning the number of solutions of a square-independent linear system in uniform subsets of $\mathbb{F}_p^n$.

**Corollary 1.4.** *For every $\epsilon > 0$ there exists $c > 0$ with the following property. Let $A$ be a subset of $\mathbb{F}_p^n$ of density $\alpha$ whose balanced function has $U^2$-norm bounded by $c$. Let $\mathcal{L} = (L_1, \ldots, L_m)$ be a square-independent system of linear forms in $d$ variables, with Cauchy-Schwarz complexity at most 2. Let $(x_1, \ldots, x_d)$ be a random element of $(\mathbb{F}_p^n)^d$. Then the probability that $L_i(x_1, \ldots, x_d) \in A$ for every $i$ differs from $\alpha^m$ by at most $\epsilon$.*

For a detailed discussion of the context of these results and their (conjectured) higher-order generalizations the reader is referred to the introduction of [GW07].
Let us now have a look at the ergodic world. Ergodic theorists are concerned with the convergence (in $L^\infty$, $L^1$ or $L^2$) of *multiple ergodic averages* of the form

$$\frac{1}{N^d} \sum_{n_1,\ldots,n_d=1}^{N} T^{p_1(n_1,\ldots,n_d)} f_1(x)\, T^{p_2(n_1,\ldots,n_d)} f_2(x)\, \ldots\, T^{p_m(n_1,\ldots,n_d)} f_m(x),$$

where $T$ is a measure preserving transformation on a probability measure space $(X, \mathcal{B}, \mu)$, the functions $f_i$ belong to $L^\infty(\mu)$ and the $p_i$ are polynomials on $\mathbb{Z}^d$. For example, the case where $d = 1$, $p_j(n) = jn$ for $j = 1, \ldots, k$ and the $f_i$ equal the indicator function $1_A$ of a set $A \in \mathcal{B}$ with $\mu(A) > 0$ appeared in Furstenberg's proof of Szemerédi's Theorem [Fur77], which states that any subset of $\mathbb{Z}$ of positive upper density contains an arithmetic progression of length $k$. More precisely, Furstenberg proved that the $\liminf_{N\to\infty}$ of the average

$$\frac{1}{N^d} \sum_{n_1,\ldots,n_d=1}^{N} \int 1_A(x) T^n 1_A(x)\, T^{2n} 1_A(x)\, \ldots\, T^{kn} 1_A(x) d\mu(x), \tag{1}$$

was strictly greater than 0. Ergodic theorists were the first to prove a multi-dimensional Szemerédi Theorem, as well as polynomial extensions [BL96] which remain beyond the reach of arithmetic combinatorics to date. However, the fact that only translation-invariant systems can be studied using such averages and, more importantly, the lack of quantitative bounds (but see [Tao06]) pose serious limitations and more than justify the search for alternative approaches via arithmetic combinatorics.

The question in ergodic theory which is analogous to the one we studied in [GW07] concerns so-called *characteristic factors* for ergodic averages of the form

$$\frac{1}{N^d}\sum_{n_1,\dots,n_d=1}^{N} T^{L_1(n_1,\dots,n_d)}f_1(x)\, T^{L_2(n_1,\dots,n_d)}f_2(x)\ \dots\ T^{L_m(n_1,\dots,n_d)}f_m(x),$$

where $T$ is a measure-preserving map on a probability measure space $(X,\mathcal{B},\mu)$, the functions $f_i$ belong to $L^\infty(\mu)$ and the $L_i$ are linear forms on $\mathbb{Z}^d$. Very roughly speaking, a characteristic factor is a system onto which one can project without losing any information about the convergence of the average under consideration. The aim is to find characteristic factors which possess enough structure to allow one to establish convergence of the above average in a rather explicit way. For example, it was shown by Host and Kra [HK05] and Ziegler [Zie07] independently that when the linear forms $L_1,\dots,L_m$ describe an arithmetic progression of length $m$, there exists a characteristic factor for the corresponding average which is isomorphic to an inverse limit of a sequence of $(m-2)$-step nilsystems. For $m=4$, these very structured objects are closely related to the quadratic factor used in [GW07], on which computations can be performed rather straightforwardly. After these remarks it should not be surprising that there is a notion of *degree* associated with a characteristic factor. What we have called the true complexity of a linear system is closely analogous to the degree of the minimal characteristic factor.

In a recent preprint [Lei07], Leibman characterizes the degree of the minimal characteristic factor for general linear as well as certain polynomial systems. Using his examples and our terminology, the system given by $\mathcal{L}_4=(x+n+m, x+2n+4m, x+3n+9m, x+4n+16m, x+5n+25m, x+6n+36m)$ has true complexity strictly greater than 1 (in fact, equal to 2), while the ever so slightly different $\mathcal{L}_5=(x+n+m, x+2n+4m, x+3n+9m, x+4n+16m, x+5n+25m, x+6n+37m)$ has true complexity 1. The crucial distinguishing factor of $\mathcal{L}_5$ is that its squares are independent, or, as Leibman puts it, that the six vectors $(1,1,1,1,1,1)$, $(1,c_1,c_2,\dots,c_5)$, $(1,d_1,d_2,\dots,d_5)$, $(1,c_1^2,c_2^2,\dots,c_5^2)$, $(1,d_1^2,d_2^2,\dots,d_5^2)$ and $(1,c_1d_1,c_2d_2,\dots,c_5d_5)$ span $\mathbb{R}^6$. (Here $c_i,d_i$ are the cofficients of $n,m$, respectively, in the linear form $i+1$. Note that the special form of the ergodic average forces one to consider translation-invariant systems only, which leads to a formulation of square-independence that is particular to systems where one variable has coefficient 1 in all linear forms.)

In his proof of Szemerédi's Theorem, Furstenberg [Fur77] developed an important tool known as the *Correspondence Principle*, which allowed him to deduce Szemerédi's combinatorial statement from the recurrence properties of a dynamical system. While the Correspondence Principle has allowed us to deduce many a combinatorial application from results in ergodic theory, our result in the $\mathbb{Z}_N$ case does not appear to follow from Leibman's result by a standard application. We shall briefly discuss this issue in the final section. For an excellent introduction to ergodic theory and its connections with additive combinatorics we refer the interested reader to [Kra06]. In this short note, we make no attempt to give a comprehensive overview of the subject but confine ourselves to describing the concepts needed to understand the parallels between [Lei07] and [GW07].

## 2. Basic Concepts in Ergodic Theory

Ergodic theory is the study of the dynamical behaviour of certain *probability measure preserving systems.*

**Definition.** *A* probability measure-preserving system *is a quadruple* $(X, \mathcal{X}, \mu, T)$ *where* $(X, \mu)$ *is a probability space and* $T : X \to X$ *is a bijective, measurable, measure-preserving transformation. This means that for all* $A \in \mathcal{X}$, $T^{-1}A \in \mathcal{X}$ *and* $\mu(T^{-1}A) = \mu(A)$.

For our purposes, we may always assume that the system $(X, \mathcal{X}, \mu, T)$ is an *ergodic* system, which means that the only sets which are left invariant under the action of $T$ have measure 0 or are in fact the whole space. This assumption is justified by a principle called *ergodic decomposition*, which says, in very rough terms, that one can decompose any measure preserving system into a number of ergodic ones. For a clear explanation see page 17 of [CFS82].

**Example 2.1.** *Let* $X = \mathbb{T}$ *be equipped with the Borel* $\sigma$*-algebra* $\mathcal{X}$ *and Haar measure* $\mu$*. Take* $T : X \to X$ *to be the rotation* $Tx = x + \alpha \bmod 1$ *for some* $\alpha \in \mathbb{R}$*. The measure preserving system* $(X, \mathcal{X}, \mu, T)$ *is ergodic if and only if* $\alpha$ *is irrational.*

The next important notion we need is that of a *factor* of a measure preserving system, that is, a subsystem which has the obvious desirable properties.

**Definition.** *A* factor *of a system* $(X, \mathcal{X}, \mu, T)$ *can be defined in several equivalent ways. Any* $T$*-invariant sub-*$\sigma$*-algebra* $\mathcal{Y}$ *of* $\mathcal{X}$ *is a factor of* $\mathcal{X}$*. A factor can also be thought of as a system* $(Y, \mathcal{Y}, \nu, S)$ *and a measurable map* $\pi : X \to Y$*, the* factor map*, such that* $\mu \circ \pi^{-1} = \nu$ *and* $S \circ \pi = \pi \circ T$ *for* $\mu$*-almost every* $x \in X$*.*

We shall be using the same letter $T$ to denote both the transformation in the original system and the transformation on the factor.

**Example 2.2.** *Let $X = \mathbb{T} \times \mathbb{T}$ be equipped with the Borel $\sigma$-algebra $\mathcal{X}$ and Haar measure $\mu$. Take $T : X \to X$ to be the transformation $T(x, y) = (x + \alpha, y + x)$ for some $\alpha \in \mathbb{R}$. Then $\mathbb{T}$ together with the rotation $x \mapsto x + \alpha$ is a factor of $X$.*

As already mentioned in the introduction, in order to study multiple ergodic averages it is useful to work on a so-called *characteristic factor.* A factor is said to be characteristic for an ergodic average if we can study the average of the projection onto the factor without losing any information about the convergence of the average. In other words, focusing on $L^2$-convergence we make the following definition.

**Definition.** *We say a factor $Y$ of $X$ is* characteristic *for the average*

$$\frac{1}{N^d} \sum_{n_1,\ldots,n_d=1}^{N} T^{p_1(n_1,\ldots,n_d)} f_1(x) T^{p_2(n_1,\ldots,n_d)} f_2(x) \ldots T^{p_m(n_1,\ldots,n_d)} f_m(x)$$

*if and only if the difference with*

$$\frac{1}{N^d} \sum_{n_1,\ldots,n_d=1}^{N} T^{p_1(n_1,\ldots,n_d)} \mathbb{E}(f_1|\mathcal{Y})(x) T^{p_2(n_1,\ldots,n_d)} \mathbb{E}(f_2|\mathcal{Y})(x) \ldots T^{p_m(n_1,\ldots,n_d)} \mathbb{E}(f_m|\mathcal{Y})(x)$$

*tends to* $0$ *in* $L^2(\mu)$*.*

Equivalently, $Y$ is characteristic for $X$ if the average converges to 0 whenever $\mathbb{E}(f_i|\mathcal{Y}) = 0$ for at least one $i = 1, 2, \ldots, m$. Here we have written $\mathbb{E}(f|\mathcal{Y})$ for the *conditional expectation* of $f$ with respect to the factor $\mathcal{Y}$, that is, the usual Hilbert space projection of $f$ onto the sub-$\sigma$-algebra $\mathcal{Y}$.

We have already mentioned that in arithmetic combinatorics, in order to show that a given linear system is governed by *some* uniformity norm all one needs is the Cauchy-Schwarz Inequality, multiple applications of which yield Theorem 1.2. We shall see that in ergodic theory, in order to show that *some* factor is characteristic for a particular average, one uses a multi-dimensional version of *Van der Corput's Lemma*, which is essentially an infinitary version of the Cauchy-Schwarz Inequality (see page 13 of [Kra06] for a standard proof which makes this obvious). We shall state Van der Corput's Lemma in dimension 3 only, for simplicity and because it is sufficient to deal with the examples we shall focus on shortly.

**Proposition 2.3.** *Suppose that $\{u_{n_1,n_2,n_3} : n_1, n_2, n_3 \in \mathbb{Z}\}$ form a bounded triple sequence of vectors in a Hilbert space. If*

$$\lim_{K\to\infty} \frac{1}{K^3} \sum_{k_1,k_2,k_3=0}^{K-1} \left| \lim_{N-M\to\infty} \frac{1}{(N-M)^3} \sum_{n_1,n_2,n_3=M}^{N-1} < u_{n_1,n_2,n_3}, u_{n_1+k_1,n_2+k_2,n_3+k_3} > \right|$$

*equals zero, then*

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

This concludes the preliminaries. In the next section we will have a closer look at how to define characteristic factors for linear systems, and collect some results about their structural properties.

## 3. Gowers Norms, Host-Kra Factors and Nilmanifolds

Recall the definition of higher-degree uniformity norms in arithmetic combinatorics, which originated in Gowers's work on Szemerédi's Theorem for longer progressions [Gow01].

**Definition.** *Let $G$ be a finite Abelian group. For any positive integer $k \geq 2$ and any function $f : G \to \mathbb{C}$, define the $U^k$-*norm *by the formula*

$$\|f\|_{U^k}^{2^k} := \mathbb{E}_{x,h_1,\dots,h_k\in G}\prod_{\boldsymbol{\omega}\in\{0,1\}^k}C^{|\boldsymbol{\omega}|}f(x+\sum_i\omega_ih_i),$$

*where $C^{|\boldsymbol{\omega}|}f = f$ if $\sum_i \omega_i$ is even and $\overline{f}$ otherwise.*

By a special case of Proposition 1.2, which was in fact proved in [Gow01], the $U^{k+1}$-norm governs the average over arithmetic progressions of length $k$ (this is because progressions of length $k$ have Cauchy-Schwarz complexity $k-2$). A family of semi-norms analogous to the $U^k$-norms have recently appeared in the work of Host and Kra [HK05].

**Definition.** *For $f \in L^\infty(\mu)$ and $k \in \mathbb{N}$., we define the* Host-Kra semi-norms *as*

$$|||f|||_k := \left(\int_{X^{[k]}} f\otimes\cdots\otimes f d\mu^{[k]}\right)^{1/k}.$$

Of course we haven't defined the measure $\mu^{[k]}$ yet, nor the space $X^{[k]}$ over which we integrate. The definition below looks rather off-putting, and we invite the reader to skip the details on first reading. However, even on more superficial inspection it can be intuited that the construction of the measure $\mu_k$ encodes the structure of combinatorial cubes of dimension $k$.

**Definition.** *Let $X^{[k]} = X^{2^k}$ and define $T^{[k]} : X^{[k]} \to X^{[k]}$ by $T^{[k]} = T \times \cdots \times T$ ($2^k$ times). We write a point $\mathbf{x} \in X^{[k]}$ as $\mathbf{x} = (x_\epsilon)_{\epsilon\in\{0,1\}^k}$ and make the natural identification of $X^{[k+1]}$ with $X^{[k]}\times X^{[k]}$, writing $\mathbf{x} = (\mathbf{x}', \mathbf{x}'')$ for a point of $X^{[k+1]}$, with $\mathbf{x}', \mathbf{x}'' \in X^{[k]}$. By induction, we define a measure $\mu^{[k]}$ on $X^{[k]}$ invariant under $T^{[k]}$. Set $\mu^{[0]} := \mu$. Let $\mathcal{I}^{[k]}$ be the invariant*

*$\sigma$-algebra of $(X^{[k]}, \mathcal{X}^{[k]}, \mu^{[k]}, T^{[k]})$. Then $\mu^{[k+1]}$ is defined to be the* relatively independent joining *of $\mu^{[k]}$ with itself over $\mathcal{I}^{[k]}$, meaning that if $F$ and $G$ are bounded functions on $X^{[k]}$,*

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

Since $(X, \mathcal{X}, \mu, T)$ is assumed to be ergodic, $\mathcal{I}^{[0]}$ is trivial and $\mu^{[1]} = \mu \times \mu$. Just like the $U^k$-norms in arithmetic combinatorics, these seminorms are nested, in the sense that they satisfy

$$|||f|||_1 \leq |||f|||_2 \leq \cdots \leq |||f|||_k \leq \cdots \leq \|f\|_\infty,$$

and a *Gowers-Cauchy-Schwarz*-type inequality holds, that is,

$$\Big| \prod_{\epsilon \in \{0,1\}^k} f_\epsilon(x_\epsilon) d\mu^{[k]} \Big| \leq \prod_{\epsilon \in \{0,1\}^k} |||f_\epsilon|||_k,$$

which can be used to show that $|||.|||_k$ is indeed a semi-norm on $L^\infty(\mu)$. Moreover, it can be checked that just like the $U^k$-norms, the semi-norms $|||.|||_k$ can be defined inductively via the formula

$$|||f|||_{k+1}^{2^{k+1}} = \int_{I_k} \mathbb{E}(f^{\otimes 2^k} | \mathcal{I}^{[k]})^2 d\mu^{[k]}.$$

Together with the *Von Neumann Ergodic Theorem*, which states that for an ergodic system $(X, \mathcal{X}, \mu, T)$ and $f \in L^2(\mu)$, the $L^2$-limit as $N$ tends to infinity of $\frac{1}{N}\sum_{n=1}^N f(T^n x)$ is the constant function $\int f d\mu$, this can be rewritten as

$$|||f|||_{k+1}^{2^{k+1}} = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} |||f \cdot T^n f|||_k^{2^k}.$$

This fact in turn is a useful ingredient in the proof of Proposition 3.1 below, which represents the analogue of Theorem 1.2 and will be discussed in more detail at the start of Section 4. We refer the keen reader to page 20 of [Kra06] for a proof in the case of arithmetic progressions.

**Proposition 3.1.** *Assume that $(X, \mathcal{X}, \mu, T)$ is ergodic and let $d, k, m \in \mathbb{N}$. Suppose $\|f_i\|_\infty \leq 1$ for all $i = 1, 2, \ldots, m$, and that the system $\mathcal{L} = (L_1, L_2, \ldots, L_m)$ in $d$ variables has Cauchy-Schwarz complexity $k$. Then*

$$\limsup_{N \to \infty} \left\| \frac{1}{N^d} \sum_{n_1, n_2, \ldots, n_d = 0}^{N-1} T^{L_1(n_1, \ldots, n_d)} f_1(x) \ \ldots \ T^{L_m(n_1, \ldots, n_d)} f_m(x) \right\|_2 \ll \min_{l=1,2,\ldots,m} |||f_l|||_{k+1}.$$

With the definitions in place, it is now straightforward to define the sequence of so-called *Host-Kra factors*, which first appeared in [HK05].

**Definition.** *Given a measure-preserving system $(X, \mathcal{X}, \mu, T)$, there is a nested sequence of factors $\mathcal{Z}_k$ of $X$ such that for any bounded function $f$ on $X$*

$$|||f|||_{k+1} = 0 \text{ if and only if } \mathbb{E}(f|\mathcal{Z}_k) = 0.$$

It follows straight from this definition combined with Proposition 3.1 that the factors $\mathcal{Z}_k$ are characteristic for systems of Cauchy-Schwarz complexity $k$. In particular, $\mathcal{Z}_1$ is characteristic for the average along 3-term progressions, while the factor $\mathcal{Z}_2$ controls 4-term progressions.

Let us pause for a moment to compare this situation with our combinatorial approach: In order to concentrate on the structured part in arithmetic combinatorics, we needed a deep $U^3$-*inverse theorem* which allowed us to decompose any bounded function into a quadratically structured and a quadratically uniform part. In ergodic theory, the fact that the factors $\mathcal{Z}_k$ are characteristic for systems of Cauchy-Schwarz complexity $k$ follows straight from the definition and Proposition 3.1. The real difficulty lies in giving a geometric description of the factors defined in this very "soft" way.

Having said that, it is not hard to see that the first factor in this sequence $\mathcal{Z}_1$ corresponds to the classical *Kronecker factor*. There are many equivalent ways of describing the Kronecker factor $\mathcal{K}$ of a measure-preserving system which do not use the semi-norm $|||.|||_2$.

- $\mathcal{K}$ is the largest abelian group rotation factor.
- $\mathcal{K}$ is the smallest sub-$\sigma$-algebra of $\mathcal{X}$ with the property that every member of $\mathcal{I}^{[1]}$ is measurable with respect to $\mathcal{K} \otimes \mathcal{K}$.
- The measure $\mu^{[2]}$ is relatively independent with respect to $\mathcal{K}^4$ and the factor $\mathcal{K}$ of $X$ is minimal with this property.

**Example 3.2.** *Let $X = \mathbb{T} \times \mathbb{T}$ be equipped with the Borel $\sigma$-algebra and Haar measure. Fix $\alpha \in \mathbb{T}$ and define $T : X \to X$ by*

$$T(x, y) = (x + \alpha, y + x)$$

*The system is ergodic if and only if $\alpha \notin \mathbb{Q}$, and it is not isomorphic to a group rotation. The Kronecker factor of $X$ is the factor $\mathbb{T}$ equipped with the rotation $x \mapsto x + \alpha$. We say $X$ is a* skew extension *of $\mathbb{T}$ by another copy of $\mathbb{T}$.*

It is not hard to see directly that $|||f|||_2$ equals the $l^4$-norm of the Fourier transform of $f$ projected onto the Kronecker factor, and that the Kronecker factor is characteristic for studying ergodic averages along 3-term progressions (see page 21 of [Kra06]). This corresponds to saying that ordinary Fourier analysis suffices in this case.

In order to study longer progressions, higher-order factors are needed. The *Conze-Lesigne factor*, which in modern terminology represents the second level in the series of Host-Kra factors, was introduced by Conze and Lesigne in a series of papers [CL84], [CL87], [CL88]. Equivalent and more explicit descriptions were given by Rudolph [Rud95] and Host and Kra [HK01], and we refer the interested reader to these works for more detail.
It turns out that every Conze-Lesigne system is the inverse limit of a sequence of 2-step nilsystems (see Theorem 18 in [HK04]). More generally, Host and Kra proved the following deep structure theorem in [HK05]:

**Theorem 3.3.** *For each integer $k$, the factor $\mathcal{Z}_k$ is isomorphic to an inverse limit of $k$-step nilsystems.*

In order to make use of this structure theorem, we need to understand what a *$k$-step nilsystem* is, as well as what it means to be an *inverse limit* of a sequence of such systems.

**Definition.** *Let $G$ be a group. If $g, h \in G$, let $[g, h] = g^{-1}h^{-1}gh$ denote the* commutator *of $g$ and $h$. If $A, B \subset G$, we write $[A, B]$ for the subgroup of $G$ spanned by $\{[a, b] : a \in A,\ b \in B\}$. The* lower central series

$$G = G_1 \supset G_2 \supset \cdots \supset G_j \supset G_{j+1} \supset \ldots$$

*of $G$ is defined by setting $G_1 = G$ and $G_{j+1} = [G, G_j]$ for $j \geq 1$. We say that $G$ is $k$-*step nilpotent *if $G_{k+1} = \{1_G\}$. If $G$ is a $k$-step nilpotent Lie group and $\Gamma$ is a discrete co-compact subgroup, the compact manifold $X = G/\Gamma$ is a $k$-*step nilmanifold*. The group $G$ acts naturally on $X$ by left translation, that is if $a \in G$ and $x \in X$, then the translation $T_a$ by $a$ is given by $T_a(x\Gamma) = (ax)\Gamma$. There is a unique Borel probability measure $\mu$ (the* Haar measure*) on $X$ that is invariant under this action. For a fixed element $a \in G$, we say that the system $(G/\Gamma, \mathcal{G}/\Gamma, T_a, \mu)$ is a $k$-*step nilsystem.

Important examples of nilsystems include the circle nilflow (Example 2.1, easily seen to be a 1-step nilsystem by setting $G = \mathbb{R}$ and $\Gamma = \mathbb{Z}$ in the above definition), the skew torus (Example 3.2, a primitive 2-step nilsystem), and the *Heisenberg nilflow*, which we shall discuss in Example 3.4 below. More information on these basic examples can be found in both [Kra06] and [GT06b].

**Example 3.4.** *Let $G$ be the Heisenberg group $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$ with multiplication given by*

$$(x, y, z) * (u, v, w) = (x + u, y + v, z + w + xv),$$

*which is a 2-step nilpotent Lie group (and is perhaps more easily thought of as the group of upper-diagonal real matrices with 1s on the diagonal). Take the discrete co-compact subgroup* $\Gamma = \mathbb{Z}\times\mathbb{Z}\times\mathbb{Z}$, *so that* $X = G/\Gamma$ *is a 2-step nilmanifold. Then the transformation* $T$ *defined as translation by* $(g_1, g_2, g_3) \in G$ *together with the Borel* $\sigma$*-algebra* $\mathcal{X}$ *and Haar measure* $\mu$ *defines a 2-step nilsystem. This system is ergodic if and only if* $g_1$ *and* $g_2$ *are rationally independent. The compact abelian group* $G/G_2\Gamma$ *is isomorphic to* $\mathbb{T}^2$*, and the rotation by* $(g_1, g_2)$ *on* $\mathbb{T}^2$ *is ergodic. This factor of* $X$ *represents the Kronecker factor* $\mathcal{Z}_1$*.*

It is not terribly important to us to know what exactly an *inverse limit* is, since it behaves well enough to always allow us to concentrate on a single nilmanifold, but for the sake of completeness we present the definition below.

**Definition.** *The system* $(X, \mathcal{X}, \mu, T)$ *is an* inverse limit *of a sequence of factors* $\{(X_j, \mathcal{X}_j, \mu_j, T)\}_{j\in\mathbb{N}}$ *if* $\{\mathcal{X}_j\}_{j\in\mathbb{N}}$ *is an increasing sequence of* $T$*-invariant sub-*$\sigma$*-algebras such that* $\bigvee_{j\in\mathbb{N}} \mathcal{X}_j = \mathcal{X}$ *up to sets of measure. If each system* $(X_j, \mathcal{X}_j, \mu_j, T)$ *is isomorphic to a* $k$*-step nilsystem, then* $(X, \mathcal{X}, \mu, T)$ *is an* inverse limit of $k$-step nilsystems.

As indicated earlier, nilmanifolds possess an enormous amount of structure, so by reducing to the study of averages on nilmanifolds via Proposition 3.1 and Theorem 3.3, many questions about the convergence of ergodic averages on abstract measure-preserving systems become explicit computations. Before we look at the general case, however, let us consider in more detail the simple 2-step nilsystem that is the skew torus.

## 4. A Square-Independent System on the Skew Torus

From now on we shall focus our attention on one of the examples of square-independent systems which was mentioned in the introduction, namely

$$\mathcal{L}_2 = (x, x+n, x+m, x+n+m, x+n-m, x+m-n).$$

It is easy to check that this linear system has Cauchy-Schwarz complexity 2 and is translation invariant. First, we shall see that the factor $\mathcal{Z}_2$ is characteristic for the average along $\mathcal{L}_2$, which is a special case of Proposition 3.1. The proof uses Van der Corput's Lemma 2.3 and the inductive definition of the semi-norm $\|\!|\cdot|\!\|_2$ given at the start of Section 3. It follows from a refined analysis of Proposition 5 in [Lei] and is left as an exercise.

**Proposition 4.1.** *Suppose $(X, \mathcal{X}, \mu, T)$ is an ergodic measure-preserving system, and let $E = \{(1,0),(0,1),(1,1),(1,-1),(-1,1)\}$. If $\|f\|_\infty \leq 1$, then*

$$\limsup_{N\to\infty} \left\| \frac{1}{N^2} \sum_{n,m=1}^{N} \prod_{\epsilon\in E} f \circ T^{\epsilon\cdot n} \right\|_{L^2(\mu)} \ll \|\!|f|\!\|_2.$$

*In other words, the factor $\mathcal{Z}_2$ is characteristic for the system $\mathcal{L}_2$.*

By Proposition 3.1 and Theorem 3.3 we are now in the fortunate position to know that we can reduce to the case where our system is a 2-step nilmanifold. Our next aim would be to show that, in fact, the Kronecker factor is the minimal characteristic factor for $\mathcal{L}_2$. For illustrative purposes we now focus on the case of the simplest possible 2-step nilsystem only, the skew torus discussed in Example 3.2 of the previous section.
Recall that the skew torus was defined by setting $X = \mathbb{T} \times \mathbb{T}$, equipped with Borel $\sigma$-algebra $\mathcal{X}$ and Haar measure $\mu$. We take $T : X \to X$ to be the transformation $T(x, y) = (x + \alpha, y + x)$ for some $\alpha \in \mathbb{R}$. This is a 2-step nilsystem, whose Kronecker factor is $\mathbb{T}$ together with the rotation $x \mapsto x+\alpha$. Iterating the transformation $T$, we find that the $n^{th}$ iterate is given by the formula

$$T^n(x, y) = (x + n\alpha, y + nx + \frac{n(n+1)}{2}\alpha).$$

It is now not difficult to compute the average explicitly. For example, suppose $f$ is a Riemann integrable function. Standard approximation arguments allow us to reduce to the case of a continuous function, and by Weierstrass approximation and linearity we are in fact justified in thinking of $f$ as a simple exponential. Inserting the formula for $T^n$ in the average

$$\frac{1}{N^2} \sum_{n,m=1}^{N} \prod_{\epsilon\in E} f \circ T^{\epsilon\cdot n}$$

and replacing each instance of $f$ by an appropriate exponential function, we find that the square-independence of $\mathcal{L}_2$ implies that there is always a non-zero quadratic coefficient of $\alpha$. This fact combined with the uniform distribution of the fractional part of $n^2\alpha$ allows us to conclude that the orbit of the diagonal $\Delta_X = \{(x, x, \ldots, x) : x \in X\}$ is uniformly distributed on the fibres over the Kronecker factor (in this case, the second co-ordinate). This in turn means that it is in fact possible to project down to the Kronecker factor without affecting the convergence of the limit of the average. Since it would not be instructive to include the full details of this computation, we leave them to the interested reader.

4.1. **A General 2-Step Nilmanifold.** The purpose of this section is to provide some intuition for the general case of a 2-step nilmanifold, and to illustrate what we mean by "parameterising" such a manifold. We shall not attempt to reproduce any proofs, but rather provide a tourist's guide to [Lei07] for the interested reader. We shall assume that we have proved Proposition 4.1 and are therefore able to restrict our attention to a 2-step nilmanifold.

Given an $s$-step nilmanifold $X = G/\Gamma$, there exists a sequence of natural factors $X = X_s \to X_{s-1} \to \ldots \to X_1 \to X_0 = \{1_X\}$ defined by $X_j = G/(\Gamma G_{j-1})$. For each $j$, $X_j$ is a $j$-step nilmanifold. This comes with a sequence of projections $\pi_j : X \to X_j$. In our case $s = 2$, so we are looking at the sequence $X_2 = X \to X_1 = G/(\Gamma G_2) \to X_0 = \{1\}$. The projection $\pi_1$ takes the simple form $G/\Gamma \to G/\Gamma G_2$.

We want to show that the factor $X_1$ is characteristic for the average along $\mathcal{L}_2$ which we were studying in the preceding section. In fact, it is possible to give a completely explicit description of the orbit of the diagonal $\Delta_X = \{(x, x, \ldots, x) : x \in X\}$ under a system of linear actions. For example, for a simple linear system of 5 forms in 2 variables such as $\mathcal{L}_2$, it can be shown that the orbit of the diagonal $\Delta_X$ is of the form $\pi^5(H)$ with $H$ a rational subgroup of the form

$$\left\langle \begin{pmatrix} b_0\, b_1^{c_1}\, b_2^{d_1}\, b_3^{c_1^2}\, b_4^{c_1 d_1}\, b_5^{d_1^2} \\ \vdots \\ b_0\, b_1^{c_5}\, b_2^{d_5}\, b_3^{c_5^2}\, b_4^{c_5 d_5}\, b_5^{d_5^2} \end{pmatrix} : b_0, b_1, b_2 \in G, b_3, b_4, b_5 \in G_2 \right\rangle,$$

where we have written $E = \{(1,0), (0,1), (1,1), (1,-1), (-1,1)\} = \{(c_i, d_i) : i = 1, 2, \ldots, 5\}$ for the coefficients of $n$ and $m$ in $\mathcal{L}_2$. This is the main content of Proposition 6.3 in [Lei07], which we have illustrated using an adaptation of Example 6.7 in that paper. But for all $i = 1, 2, \ldots, 5$, we can now rewrite

$$b_0\, b_1^{c_i}\, b_2^{d_i}\, b_3^{c_i^2}\, b_4^{c_i d_i}\, b_5^{d_i^2}$$

as a product

$$b_0\, b_1^{c_i}\, b_2^{d_i}\, (a_2^{-d_i}\, a_1^{-c_i}\, a_0^{-1}\, a_0\, a_1^{c_i}\, a_2^{d_2})\, b_3^{c_i^2}\, b_4^{c_i d_i}\, b_5^{d_i^2}$$

with $a_0, a_i, a_2 \in G_2$, which in turn can be expressed as

$$(a_0^{-1} b_0)(a_1^{-1} b_1)^{c_i}(a_2^{-1} b_2)^{d_i} a_0\, a_1^{c_i}\, a_2^{d_i}\, b_3^{c_i^2}\, b_4^{c_i d_i}\, b_5^{d_i^2}.$$

This reparametrisation takes place in Corollary 5.8 of [Lei07]. Finally, we know that because the system $\mathcal{L}_2$ is square-independent, the matrix of coefficients

$$\begin{pmatrix} 1 & c_1 & d_1 & c_1^2 & c_1d_1 & d_1^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & c_5 & d_5 & c_5^2 & c_5d_5 & d_5^2 \end{pmatrix}$$

has full rank, and hence the rational subgroup $H$ takes the form

$$\left\langle \begin{pmatrix} b_0\ b_1^{c_1}\ b_2^{d_1} \\ \vdots \\ b_0\ b_1^{c_5}\ b_2^{d_5} \end{pmatrix} : b_0, b_1, b_2 \in G \right\rangle \cdot G_2^5.$$

Since the projection $\pi_1 : X \to X_1$ amounted to nothing more than quotienting out by the commutator subgroup $G_2$, we see that in fact the factor $X_1$ is characteristic for a square-independent average.
Note that a very similar parametrisation can be carried out for polynomial orbits, details of which can be found in the later sections of [Lei07].

4.2. **The Correspondence Principle.** It is not clear whether Leibman's ergodic theoretic result has any number theoretic consequences of the form we saw in Corollary forsets. In general, one uses the following standard tool for transfering ergodic theoretic to combinatorial statements, which originated in Furstenberg's proof of Szemerédi's Theorem [Fur77] and is now known as the *Correspondence Principle.*

**Proposition 4.2.** *Let $E$ be a set of integers of positive upper density. Then there exist an ergodic system $(X, \mathcal{X}, \mu, T)$ and a set $A \in \mathcal{X}$ with $\mu(A) = d^*(E)$ such that*

$$\mu(T^{m_1}A \cap \cdots \cap T^{m_k}A) \leq d^*((E + m_1) \cap \cdots \cap (E + m_k))$$

*for all integers $k \geq 1$ and all $m_1, ..., m_k \in \mathbb{Z}$.*

While it is easily seen that this proposition implies Szemerédi's Theorem for progressions of length $k$ once a positive limit for the ergodic average (1) is established, when one attempts to transfer Leibman's result to a statement such as Corollary forsets, one only obtains a lower bound on the number of solutions rather than an asymptotically exact statement.

4.3. **Remarks.** Leibman [Lei07] is able to determine the true complexity of all translation-invariant linear systems, not just those of Cauchy-Schwarz complexity 2. The main reason for this level of generality is that Host and Kra's structure theorem (Theorem 3.3) is available for all $k$, unlike the situation in arithmetic combinatorics where the decomposition theorem depends on the existence of a suitable $U^k$-inverse theorem, which has only been proved for $k \leq 3$. The fact that ergodic theorists are able to deal with polynomial systems is another point of envy. Indeed, it turns out that the seminorms $|||.|||_k$ also control polynomial averages when combined with *PET induction* (a linearization method which originated in [BL96]). In the finite combinatorial world, on the other hand, so-called "local" $U^k$-norms will be required in order to control polynomial averages. The reason for this is that when we consider polynomials such as $x + n^2$ inside an interval $1, 2, \ldots, N$, we are forced to restrict the range of the parameter $n$ to $\sqrt{N}$. These local uniformity norms are currently much less well understood, but see [TZ06] for more details on the emerging theory of local uniformity.

**Acknowledgements.** The author would like to thank Bryna Kra for valuable comments and discussions.

## References

[BL96] V. Bergelson and A. Leibman. Polynomial extensions of Van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.*, 9:725–753, 1996.

[CFS82] I.P. Cornfeld, S.V. Fomin, and Ya. G. Sinaĭ. *Ergodic theory*, volume 245 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1982.

[CL84] J.-P. Conze and E. Lesigne. Théorèmes ergodiques pour des mesures diagonales. *Bull. Soc. Math. France*, 112:143–175, 1984.

[CL87] J.-P. Conze and E. Lesigne. Sur un théorème ergodique pour des mesures diagonales. *Publications de l'Institut de Recherche de Mathématiques de Rennes, Probabilités*, 1987.

[CL88] J.-P. Conze and E. Lesigne. Sur un théorème ergodique pour des mesures diagonales. *C. R. Acad. Sci. Paris, Série I*, 306:491–493, 1988.

[Fur77] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.*, 31:204–256, 1977.

[Gow01] W.T. Gowers. A new proof of Szemerédi's theorem. *GAFA*, 11:465–588, 2001.

[GT06a] B.J. Green and T. Tao. Linear equations in primes. Submitted. Available at arXiv:math.NT/0606088, 2006.

[GT06b] B.J. Green and T. Tao. Quadratic uniformity of the Möbius function. Available at arXiv:math.NT/0606087, 2006.

[GW07] W.T. Gowers and J. Wolf. The true complexity of a system of linear equations. Available at arXiv:math.NT/, 2007.

[HK01] B. Host and B. Kra. Convergence of conze-lesigne averages. *Erg. Th. & Dyn. Sys.*, 21:493–509, 2001.

[HK04] B. Host and B. Kra. Averaging along cubes. In *Dynamical systems and related topics*. Cambridge University Press, Cambridge, 2004.

[HK05] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Annals of Math.*, 161(1):397–488, 2005.

[Kra06] B. Kra. Ergodic methods in additive combinatorics. Available at arXiv:math.DS/0608105, 2006.

[Lei] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables.

[Lei07] A. Leibman. Orbit of the diagonal of a power of a nilmanifold. Available at http://www.math.ohio-state.edu/∼Leibman/preprints, 2007.

[Rud95] D.J. Rudolph. Eigenfunctions of T×S and the Conze-Lesigne algebra. In *Ergodic theory and its Connections with Harmonic Analysis*, pages 369–432. Cambridge Univ. Press, New York, 1995.

[Tao06] T. Tao. A quantitative ergodic theory proof of Szemerédi's theorem. Available at arXiv:math.CO/0405251, 2006.

[TZ06] T. Tao and T. Ziegler. The primes contain arbitrarily long polynomial progressions. arXiv:math/0610050, 2006.

[Zie07] T. Ziegler. Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.*, 20:53–97, 2007.

Department of Pure Mathematics and Mathematical Statistics, Wilberforce Road, Cambridge CB3 0WB, U.K.

*E-mail address*: `J.Wolf@dpmms.cam.ac.uk`