PHD thesis

Multiple ergodic averages along sequences of polynomial growth

A thesis submitted for the degree of Doctor of Philosophy by

KONSTANTINOS TSINAS

supervised by
PROF. NIKOS FRANTZIKINAKIS

Department of Mathematics and Applied Mathematics,
University of Crete

June 2024

Contents

Abstract

Ergodic theory has been an active area of research in recent decades. Furstenberg's original work in the proof of Szemerédi's theorem was the spark for the development of a whole new research field, since dynamical methods were then understood to be a potent tool for solving combinatorial problems. These methods naturally led researchers to ask many follow-up questions and nowadays we have generalizations of Furstenberg's results concerning patterns beyond arithmetic progressions as well as deep theorems describing the structure of measure-preserving systems.

In this thesis, we investigate the problem of convergence of multiple ergodic averages along sequences that arise from functions that have polynomial growth and some extra regularity properties, such as monotonicity and smoothness. Typical examples include the polynomials or the fractional powers n^c, where $c > 0$ is not an integer. We show that under some simple assumptions on the growth rates of the functions we have convergence of multiple ergodic averages along these sequences in all measure-preserving systems. As a consequence of these results, we derive several combinatorial applications showing that all subsets of $\mathbb{Z}$ with positive density contain patterns of a specific form. In the case of nilmanifolds, we prove pointwise convergence results for these averages and then use well-known structure theorems to deduce convergence results for general measure-preserving systems. Furthermore, we ask the same questions for multiple ergodic averages evaluated along the prime numbers and we show that under the same assumptions, the corresponding averages converge and the limit is the same as the limit of the typical averages along the naturals.

The results of this thesis are contained in the following articles (listed in chronological order):

1) K. Tsinas. Joint ergodicity of Hardy field sequences. *Transactions of the American Mathematical* Society, **376**:3191–3263, 2023.

2) K. Tsinas. Pointwise convergence in nilmanifolds along smooth functions of polynomial growth. *Ergodic Theory and Dynamical Systems.* Published online p:1-46. doi:10.1017/etds.2023.6, 2023

3) A. Koutsogiannis and K. Tsinas. Ergodic averages for sparse sequences along primes. Preprint 2023, arXiv.2309.0493

Acknowledgements

First of all, I would like to thank my advisor Nikos Frantzikinakis for his constant encouragement, his guidance and his enthusiasm to help during all these years. To this day, I always feel like I am learning something new every time I discuss mathematical ideas with him. It is not an understatement to say that this thesis would never have been completed without him motivating, guiding and teaching me along the way.

I would like to thank Andreas Koutsogiannis for his aid throughout all these years as a teacher, as a research collaborator and as a friend. All results obtained in Section 1.3 were obtained in a collaborative article with him. I would also like to thank my friend Borys Kuca, especially for his advice and the many fruitful mathematical discussions that we had the year that we were both in Crete. I would also like to thank Silouanos Brazitikos, whom I have known for many years, for all his help and advice. I am grateful to the late professor Dimitris Gatzouras who was the one to introduce me to the beautiful world of ergodic theory during my Master's studies. I also want to thank professor Apostolos Giannopoulos for his help during my Bachelor and Master's studies and also for introducing me to my eventual advisor during my Master's thesis. I would like to thank professors Giorgos Costakis and Mihalis Kolountzakis for serving in my dissertation commitee.

Last and most important of all, I would like to thank my family for always being supportive and encouraging throughout my journey all these years.

This thesis was supported by the Hellenic Institute for Research and Innovation initially by the Research Grant - ELIDEK HFRI-FM17-1684 and then by ELIDEK-Fellowship number 5367 (3rd Call for HFRI Ph.D. Fellowships). The author gratefully acknowledges their contribution.

Chapter 1

Introduction

In this thesis, we investigate problems concerning the convergence of multiple averages along many sequences of interest and we derive several combinatorial applications, indicating that positive density subsets of the integers contain many patterns of a specific form. The idea of using tools from ergodic theory and dynamical systems can be traced back to Furstenberg [21] who gave a different proof of the following theorem of Szemerédi [53].

Theorem A (Szemerédi's theorem)**.** *Any positive density subset of $\mathbb{N}$ contains arbitrarily long arithmetic progressions.*

In his seminal work [21], Furstenberg established his multiple recurrence theorem for general measure-preserving systems, which, in conjunction with Furstenberg's correspondence principle, yields another proof of Theorem A. In the field of ergodic theory, a *measure-preserving system* is a probability space $(X, \mathcal{X}, \mu)$ equipped with an invertible *measure-preserving transformation* T. This means that $\mu(A) = \mu(T^{-1}A)$ for any measurable set A. One can also consider a probability space with many commuting transformations measure-preserving transformations $T_1, \dots, T_k$ acting on it at the same time. We then call $(X, \mathcal{X}, \mu, T_1, \dots, T_k)$ a *measure-preserving system with k commuting transformations*. In several places throughout the text, we will omit the σ-algebra from the notation if there is no confusion.

First of all, we state the correspondence principle. Assume $E \subset \mathbb{N}$. Then, we define the upper density of the set E as the limit

$$\bar{d}(E) := \limsup_{N\to\infty} \frac{|E \cap [1, N]|}{N}$$

and the lower density $\underline{d}$ is defined similarly with lim inf instead of lim sup. If these limits coincide, then we say that the set E has natural density $d(E)$ equal to the limit.

Theorem B (Furstenberg's correspondence principle)**.** *For any set $E \subset \mathbb{N}$ with positive upper density, there exist an invertible measure preserving system $(X, \mathcal{X}, \mu, T)$ and a measurable set $A \subset X$, such that $\bar{d}(E) = \mu(A)$ and for any $r_1, ..., r_k \in \mathbb{Z}$, we have*

$$\bar{d}(E \cap (E - r_1) \cap \cdots \cap (E - r_k)) \geq \mu(A \cap T^{-r_1}A \cap \cdots \cap T^{-r_k}A).$$

Theorem C (Furstenberg's multiple recurrence theorem)**.** *Let k be a positive integer, $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system and let $A \in \mathcal{X}$ be a set with positive measure. Then, we have*

$$\liminf_{N\to+\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A) > 0.$$

In particular, there exists $n \in \mathbb{N}$ for which $\mu(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A) > 0$.

Furstenberg's correspondence principle allows us to translate problems of finding structures in subsets of $\mathbb{N}$ to multiple recurrence problems in dynamical systems, where new analytic tools are available to tackle the task at hand. Typically, we also want to determine the more difficult question of whether limits such as those in Theorem C exist or not and, if possible, find an explicit expression for them. In general, our problems involve the study of the multiple ergodic averages

$$\frac{1}{N}\sum_{n=1}^{N} f_1(T^{a_1(n)}x)\cdot ... \cdot f_k(T^{a_k(n)}x) \tag{1.1}$$

where $a_1(n), ..., a_k(n)$ are sequences of integers and T is an invertible measure preserving map on a probability space $(X, \mathcal{X}, \mu)$. In this thesis, we will only focus on the limit in the L^2 sense. Similar questions can be asked about pointwise convergence of these averages, although the theory and the results in pointwise convergence are still far behind the knowledge we have on the L^2 theory. Furthermore, understanding the L^2 limiting behavior of the previous averages is sufficient for the combinatorial applications in almost all cases.

In the case of Theorem C, one has to study the limiting behavior of the averages

$$\frac{1}{N}\sum_{n=1}^{N} T^n f_1 \cdot T^{2n} f_2 \cdot \ldots \cdot T^{kn} f_k \tag{1.2}$$

and then specialize to the case $f_1 = f_2 = \cdots = f_k = \mathbf{1}_A$.

Furstenberg's multiple recurrence is sufficient to derive Theorem A, although Furstenberg's original work did not determine whether the associated average converges or not. The crux of Furstenberg's argument was his deep structure theorem, showing how an ergodic system can be constructed by the trivial system through a series of *extensions* (typically, infinitely many) with particular properties. Using this theorem, one then has to verify that if a system possesses the multiple recurrence property of Theorem C, then any of the extensions like the ones mentioned above will produce a new system that still has the multiple recurrence property. Therefore, the multiple recurrence theorem can be lifted from the trivial system to any general ergodic system and, then, to all measure preserving systems using a standard ergodic decomposition argument.

Relying on Furstenberg's structure theorem, many authors provided generalizations of Theorem C even to systems involving several commuting transformations acting on the same probability space. In particular, Bergelson and Leibman [3] generalized Theorem C (building on work of Furstenberg and Katznelson [22] in the case of linear iterates) to polynomials of higher degree.

Theorem D (Bergelson-Leibman theorem)**.** *Let $k \in \mathbb{N}$, $(X, \mathcal{X}, \mu, T_1, \ldots, T_k)$ be a measure-preserving system with k commuting measure-preserving transformations and let $p_1, \ldots, p_k$ be integer polynomials with zero constant term. Then, there exists a positive integer n, such that*

$$\mu(A \cap T_1^{-p_1(n)}A \cap \cdots \cap T_k^{-p_k(n)}A) > 0.$$

Once again, Furstenberg's correspondence principle allows someone to prove that positive density subsets of $\mathbb{Z}$ contain patterns of the form $(m, m + p_1(n), \ldots, m + p_k(n))$. Actually, the fact that we can allow many distinct transformations T_i yields combinatorial results for subsets of $\mathbb{Z}^k$.

In 2005, Host and Kra established their famous structure theorem [29]. They show that, for any $k \in \mathbb{N}$, each ergodic system is an extension of a special system, called the "factor of order k" (denoted by $\mathcal{Z}_k$), which in turn is the inverse limit of systems with special algebraic structure, called nilsystems. The upside of this result is that many problems concerning averages such as (1.1) can be reduced through a series of analytic methods to determining the limiting behavior only in the case of these special factors, whose structure is now well understood. For instance, the Host-Kra structure theory can be used to show convergence of the averages in (1.2) and this was carried out in [29]. Subsequently, an independent proof of this was given by Ziegler [60]. This theorem has also proven very effective in handling the case of polynomial iterates. In particular, Leibman [38] proved that the averages in (1.1) converge in $L^2(\mu)$ in the case where $a_i(n)$ are integer polynomials. In this setting, the structure factors are called characteristic for the polynomial averages. More specifically, we say that a factor $(Y, \mathcal{Y}, \nu, T)$ [1] of the system $(X, \mathcal{X}, \mu, T)$ is characteristic for the averages (1.1), if we have that

$$\Big\|\frac{1}{N}\sum_{n=1}^{N} T^{a_1(n)}f_1 \cdot ... \cdot T^{a_k(n)}f_k - \frac{1}{N}\sum_{n=1}^{N} T^{a_1(n)}\mathcal{E}_\mu(f_1|Y)\cdot ... \cdot T^{a_k(n)}\mathcal{E}_\mu(f_k|Y)\Big\|_{L^2(\mu)} \to 0$$

[1] For a strict definition of a factor system, we refer the reader to the following chapter.

as $N \to +\infty$. In order to make sense of the conditional expectation with respect to a factor, we remark briefly that there is a correspondence between factors of a system and T-invariant sub σ-algebras (cf. Chapter 2).

The main topic of this thesis is to investigate convergence results along sequences with polynomial-like behaviour. Some typical examples of sequences that we study are polynomial sequences with real coefficients or sequences involving fractional powers $\lfloor n^c \rfloor$. More generally, we will consider sequences of the form $\lfloor f(n) \rfloor$ where f is a function that has polynomial growth[2] and several more regularity properties (like, for example, smoothness and monotonicity). Furthermore, we will tackle similar problems for the same sequences evaluated along the prime numbers. It is generally understood that the most natural class of functions to study are the functions that belong to a *Hardy field*.

Let $\mathcal{B}$ denote the set of germs at infinity of real valued functions defined on a half-line $[x, +\infty]$. That means two functions f, g will be considered the same if $f(t) = g(t)$ for t sufficiently large. Then, $(\mathcal{B}, +, \cdot)$ is a ring.

Definition 1.0.1. *A sub-field $\mathcal{H}$ of $\mathcal{B}$ that is closed under differentiation is called a Hardy field. We will say that $a(n)$ is a Hardy sequence, if for $n \in \mathbb{N}$ large enough we have $a(n) = f(n)$ for some function $f \in \mathcal{H}$.*

We will make some small abuse of language and sometimes also refer to sequences of the form $\lfloor f(n) \rfloor$ as Hardy sequences, since the rounding functions will appear several times throughout the text.

An example of a Hardy field is the field $\mathcal{LE}$ of logarithmico-exponential functions. These are defined on a half line of $\mathbb{R}$ by a finite combination of the operations $+, -, \cdot, \div$, exp, log and composition of functions acting on a real variable t and real constants. This class contains the polynomials $p(t)$, the fractional powers t^c, as well as functions like $t \log t$, $t^{(\log t)^2}$ and $e^{\sqrt{t}}/t^2$.

The problem of determining whether or not multiple ergodic averages with Hardy field iterates converge has been studied extensively. The simpler case involving one function is well understood through the work of Boshernitzan, Kolesnik, Quas and Wierdl [7], relying in particular on Boshernitzan's characterization of when a Hardy sequence is equidistributed modulo 1 [5]. To be more precise, we have the following theorems.

Theorem E (Boshernitzan)**.** *Let the function $a \in \mathcal{H}$ have polynomial growth. Then, the sequence $a(n)$ is equidistributed* mod 1 *if and only if*

$$\lim_{t \to +\infty} \frac{|a(t) - q(t)|}{\log t} = +\infty \quad \textit{for any polynomial} \quad q(t) \in \mathbb{Q}[t]. \tag{P}$$

This theorem characterizes when a Hardy sequence is equidistributed modulo 1. One can combine this with the spectral theorem for unitary operators and (almost) determine whether the averages in (1.1) converge in the case of $k = 1$ and when a is not "essentially" equal to a polynomial. The polynomial case was also well understood before the investigation of Hardy sequences began and, as it turns out, there is also one final extra case where convergence still holds. All of the above are the content of the following theorem, which was proven in [7] by Boshernitzan, Kolesnik, Quas and Wierdl.

Theorem F. *Let $\mathcal{H}$ be a Hardy field and $a \in \mathcal{H}$ be a function of polynomial growth satisfying one of the following three conditions:*
i) we have

$$\lim_{t \to +\infty} \frac{|a(t) - cq(t)|}{\log t} = +\infty \textit{ for all } c \in \mathbb{R} \textit{ and } q(t) \in \mathbb{Z}[t],$$

ii) there exist $c, d \in \mathbb{R}$ and $q(t) \in \mathbb{Z}[t]$ such that

$$\lim_{t \to +\infty} |a(t) - cq(t)| = d \textit{ or}$$

[2] A function f is said to have polynomial growth, if there exists a positive integer d, such that the ratio $\frac{f(t)}{t^d}$ converges to 0, as $t \to +\infty$.

iii) there exists an integer m and a positive constant C such that

$$\left|a(t)-\frac{t}{m}\right|\leq C\log t \quad \textit{for all } t\in\mathbb{R}^+.$$

Then, for any measure preserving system $(X,\mathcal{X},\mu,T)$ and function $f\in L^2(\mu)$, the averages

$$\frac{1}{N}\sum_{n=1}^{N}T^{\lfloor a(n)\rfloor}f$$

converge in $L^2(\mu)$. Furthermore, in cases i) and iii), the limit of the averages is equal to the conditional expectation $\mathcal{E}_\mu(f|I(T))$, where $I(T)$ is the invariant factor of the system.

If the system is ergodic, then the invariant factor is trivial and the conditional expectation above is constant and equal to the integral of f. If a function $a\in\mathcal{H}$ has polynomial growth and does not satisfy one of the previous three conditions, then we can show that convergence fails for at least one system.

The previous theorem (as well as Theorem E) implies that the distance of a from real multiples of integer polynomials is precisely what determines whether the Hardy sequence is good for convergence in the single iterate case. Namely, the Hardy sequence must be either far away from polynomials (i.e their distance "growing" faster than $\log t$) or very close to polynomials (i.e the distance stays bounded). The polynomials of the form $\frac{t}{m}$ for $m\in\mathbb{Z}\setminus\{0\}$ are the only exception to this, since condition iii) allows the distance to grow slower than $\log t$, but still go to infinity.

In the setting of several iterates, Frantzikinakis proved [13] that for a function $a\in\mathcal{H}$ of polynomial growth that satisfies one of conditions i), ii) or iii), the averages

$$\frac{1}{N}\sum_{n=1}^{N}T^{\lfloor a(n)\rfloor}f_1\dots T^{k\lfloor a(n)\rfloor}f_k$$

converge in $L^2(\mu)$. Using this theorem, he was able to prove a strengthening of Szemerédi's theorem, namely that positive density subsets of $\mathbb{N}$ contain arbitrarily long arithmetic progressions, where the step of the progression has the form $\lfloor a(n)\rfloor$. In the same article, Frantzikinakis showed that if the Hardy field functions $a_1,\dots,a_k$ satisfy certain growth assumptions, then the averages

$$\frac{1}{N}\sum_{n=1}^{N}T^{\lfloor a_1(n)\rfloor}f_1\dots T^{\lfloor a_k(n)\rfloor}f_k$$

converge to the product of the integrals in ergodic systems. In addition, he posed a conjecture that the same result holds under the more general assumption that all linear combinations of the functions $a_1,\dots,a_k$ satisfy condition i) in Theorem F. This condition is natural, in the sense that under this assumption and Boshernitzan's equidistribution theorem [5], we can verify this statement for the simplest systems, namely rotations on finite-dimensional tori. One of our results below verifies this conjecture.

Our results are separated into three categories. In the first case, we study the behavior of multiple ergodic averages for collections of Hardy field functions for which we expect convergence to the product of the integrals. In the following section, we will have to tackle the case where we have convergence to a limit other than the product of the integrals. In this case, we will rely on the Host-Kra structure theorem and, therefore, we have to prove convergence for nilsystems, which are the building blocks in the structure theory. We remark here briefly that we will avoid the study of nilsystems for the results in Section 1.1 thanks to a recent characterization of "joint-ergodicity" due to Frantzikinakis [17]. In the third section of this chapter, we will investigate the previous questions again in the setting where the sequences are evaluated along the sequence of prime numbers. We shall provide combinatorial applications of our ergodic theoretic results as well and we will discuss previous results in the literature in each section below.

1.1 Convergence to the product of the integrals

In order to prove the theorems in this section, we will have to work with Hardy fields $\mathcal{H}$ that contain the Hardy field $\mathcal{LE}$ of logarithmico-exponential functions and which are closed under composition and compositional inversion of functions, when defined. An example of a Hardy field that satisfies the above property is the Hardy field of Pfaffian functions. All the subsequent results in this section will be stated under the above assumption. More background on Hardy fields will be presented in the next chapter, where we also present the definition of Pfaffian functions as well. We denote by $C\mathbb{Z}[t]$ the collection of all real multiples of integer polynomials on some variable t. If $a_1, ..., a_k$ are general sequences or functions, we will denote by $\mathcal{L}(a_1, ..., a_k) \subseteq \mathcal{H}$ the set of non-trivial linear combinations of the functions $a_1, ..., a_k$ (here $\mathcal{H}$ is a vector space over $\mathbb{R}$).

Theorem 1.1.1. *[54, Theorem 1.2] Let $\mathcal{H}$ be a Hardy field that contains $\mathcal{LE}$ and is closed under composition and compositional inversion of functions, when defined [3]. Assume $a_1, ..., a_k \in \mathcal{H}$ have polynomial growth and that every function in $\mathcal{L}(a_1, ..., a_k)$ satisfies*

$$\lim_{t\to+\infty} \frac{|a(t) - q(t)|}{\log t} = +\infty \text{ for all } q(t) \in C\mathbb{Z}[t] \tag{1.3}$$

Then, for any ergodic measure preserving system $(X, \mathcal{X}, \mu, T)$ and functions $f_1, ..., f_k \in L^\infty(\mu)$, the averages

$$\frac{1}{N}\sum_{n=1}^{N} T^{\lfloor a_1(n)\rfloor} f_1 \cdot ... \cdot T^{\lfloor a_k(n)\rfloor} f_k \tag{1.4}$$

converge in mean to the product of the integrals $\int f_1 \, d\mu \cdots \int f_k \, d\mu$.

Remark. The condition on the linear combinations of the functions $a_1, ..., a_k$ can be substituted by the following more general assumption: for any real numbers $t_1, ..., t_k \in [0, 1)$, not all of them zero, we have

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} e(t_1\lfloor a_1(n)\rfloor + \cdots + t_k\lfloor a_k(n)\rfloor) = 0,$$

where we denote $e(t) = e^{2\pi i t}$ (see our notational conventions). This is a necessary and sufficient condition in order to have convergence to the product of the integrals in every ergodic system. However, it is difficult to relate this condition with the form the functions must take in Theorem F.

If we do not impose an ergodicity assumption on the system (X, μ, T), then we can show that the averages in the above theorem converge to the product

$$\mathcal{E}_\mu(f_1|\mathcal{I}_T) \cdot ... \cdot \mathcal{E}_\mu(f_k|\mathcal{I}_T),$$

where $\mathcal{E}_\mu(f|\mathcal{I}_T)$ is again the projection of f to the invariant factor of the system. This follows from an ergodic decomposition argument.

This theorem generalizes several results in the literature. In the case of real polynomials, Theorem 1.1.1 was established in [34]. As we briefly mentioned in the introduction, this result was established in [13] in the case where all functions $a_1, ..., a_k$ have different growth rates and satisfy $t^{N_i+\varepsilon} \ll a_i(t) \prec t^{N_i+1}$ for non-negative integers N_i and some $\varepsilon > 0$. More recently, 1.1.1 was established in [4] under a linear independence condition on the functions $a_1, ..., a_k$ and on all of their derivatives. It was proven, however, that if we use a weaker averaging scheme than Cesáro averages, we can establish convergence results for the corresponding multiple ergodic averages. Finally, Theorem 1.1.1 was established recently for linear combinations of tempered functions from a Hardy field and real polynomials in [17] (for functions f belonging to $\mathcal{H}$, the tempered condition is equivalent to the relation $t^k \log t \prec f(t) \ll t^{k+1}$, for some non-negative integer k). A case that is not covered in any of the previous results is the pair $(t \log t, t^2 \log t)$.

A variant of Theorem 1.1.1 for commuting transformations was proven in [14] under more restrictive conditions, which we will review when we discuss the convergence results along primes. Our methods

[3]This means that if $f, g \in \mathcal{H}$ are such that $g(t) \to +\infty$, then $f \circ g \in \mathcal{H}$ and $g^{-1} \in \mathcal{H}$.

fail to extend Theorem 1.1.1 to this case, the main reason being that we cannot establish the required seminorm estimates as in the single transformation case.

If our only objective is to determine characteristic factors for our averages, we can relax the conditions of Theorem 1.1.1 considerably. More precisely, we have the following theorem which appeared as a conjecture in [13, Problem 3]. The notion of the Host-Kra factor of a system is somewhat technical and we postpone it until the following chapter.

Theorem 1.1.2. *[54, Theorem 1.3] Let $\mathcal{H}$ be a Hardy field that contains $\mathcal{LE}$ and is closed under composition and compositional inversion of functions. Assume that the functions $a_1, ..., a_k \in \mathcal{H}$ have polynomial growth and satisfy*

$$\lim_{t\to+\infty} \frac{|a_i(t)|}{\log t} = +\infty \quad \text{for all} \ \ 1 \leq i \leq k$$

and

$$\lim_{t\to+\infty} \frac{|a_i(t) - a_j(t)|}{\log t} = +\infty \quad \text{for all} \quad i \neq j.$$

Then, there exists a positive integer s such that, for any measure preserving system $(X, \mathcal{X}, \mu, T)$, we have

$$\lim_{N\to\infty} \Big\| \frac{1}{N} \sum_{n=1}^{N} T^{\lfloor a_1(n) \rfloor} f_1 \cdot ... \cdot T^{\lfloor a_k(n) \rfloor} f_k - \frac{1}{N} \sum_{n=1}^{N} T^{\lfloor a_1(n) \rfloor} \widetilde{f}_1 \cdot ... \cdot T^{\lfloor a_k(n) \rfloor} \widetilde{f}_k \Big\|_{L^2(\mu)} = 0,$$

where $\widetilde{f}_i := \mathcal{E}_\mu(f_i | Z_s(X))$ is the projection of f_i to the s-step Host-Kra factor of the system.

Most of the work will be allocated into proving this theorem. The conditions in Theorem 1.1.1 arise from the joint ergodicity criterion of Frantzikinakis, which requires certain trigonometric averages to vanish. The conditions in the last theorem are necessary (one can consider some weakly-mixing systems that are not strongly-mixing to see this). Since for weak-mixing systems, the Host-Kra factors of any order are trivial, we get the following corollary, which extends the results in [1, Theorem 1.2] where the iterates are polynomials taking integer values on the integers, as well as some of the results in [2] involving tempered functions.

Corollary 1.1.3. *[54, Corollary 1.4]Let $\mathcal{H}$ be a Hardy field that contains $\mathcal{LE}$ and is closed under composition and compositional inversion of functions. Assume that the functions $a_1, ..., a_k \in \mathcal{H}$ have polynomial growth and satisfy*

$$\lim_{t\to+\infty} \frac{|a_i(t)|}{\log t} = +\infty \quad \text{for all} \ \ 1 \leq i \leq k$$

and

$$\lim_{t\to+\infty} \frac{|a_i(t) - a_j(t)|}{\log t} = +\infty \quad \text{for all} \quad i \neq j.$$

Then, for any weak-mixing system $(X, \mathcal{X}, \mu, T)$, we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^{\lfloor a_1(n) \rfloor} f_1 \cdot ... \cdot T^{\lfloor a_k(n) \rfloor} f_k = \int f_1 \, d\mu \cdot ... \cdot \int f_k \, d\mu,$$

where convergence takes place in $L^2(\mu)$.

Finally, we provide some combinatorial applications of our convergence results which follow by using Furstenberg's correspondence principle. As a corollary of Theorem 1.1.1, we get the following multiple recurrence result.

Corollary 1.1.4. *[54, Corollary 1.5] Suppose $\mathcal{H}$ and $a_1, \ldots, a_k$ satisfy the hypotheses of Theorem 1.1.1. Then, for any measure preserving system $(X, \mathcal{X}, \mu, T)$ and any set $A \subset X$ with $\mu(A) > 0$, we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-\lfloor a_1(n) \rfloor} A \cap \cdots \cap T^{-\lfloor a_k(n) \rfloor} A) \geq \mu(A)^{k+1}.$$

A similar result was established in [4] with lim sup in place of the limit, but under more general conditions on the functions $a_1, ..., a_k$.

Utilizing Furstenberg's correspondence principle, we can deduce a combinatorial result about large sets of integers.

Corollary 1.1.5. *[54, Corollary 1.6] Let $E \subset \mathbb{N}$ have positive upper density and suppose $\mathcal{H}$ and $a_1, \ldots, a_k$ satisfy the hypotheses of Theorem 1.1.1. Then,*

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \bar{d}(E \cap (E - \lfloor a_1(n)\rfloor) \cap \cdots \cap (E - \lfloor a_k(n)\rfloor)) \geq (\bar{d}(E))^{k+1}.$$

As we mentioned above, a similar result was obtained in [4], which was sufficient to deduce that there exist infinitely many $n \in \mathbb{N}$ for which $\bar{d}(E \cap (E - \lfloor a_1(n)\rfloor) \cap \cdots \cap (E - \lfloor a_k(n)\rfloor)) > 0$.

1.2 Pointwise convergence in nilmanifolds and the case of linearly dependent iterates

As we discussed in the introduction, a powerful theorem in the study of the averages in (1.1) is the structure theorem of Host-Kra [29], which can reduce the above problem to studying rotations on particular spaces called nilmanifolds. A *nilmanifold* is a homogeneous space $X = G/\Gamma$, where G is a nilpotent Lie group and Γ is a discrete cocompact subgroup. In this sense, nilmanifolds are the nilpotent analog of finite dimensional tori, which constitute the abelian case of the nilpotent theory. In this section, our results will focus on determining the distribution of orbits in a nilmanifold along Hardy sequences.

The reader may observe that there is no mention of nilmanifolds in the results of the previous section. The reason is that we can bypass the use of the Host-Kra structure theorem by using the joint ergodicity criterion of Frantzikinakis. In simplistic terms, this joint ergodicity criterion implies that in the case of convergence to the product of the integrals, we only need to establish convergence for "abelian" nilmanifolds, which are precisely the finite-dimensional tori. In this case, our linear independence assumption (1.3) and Theorem E allow for a quick verification of this condition. However, the joint ergodicity criterion cannot be used when there are linear dependencies between the Hardy field functions and, therefore, we need to investigate the non-abelian nilmanifolds in this case.

In order to prove the corresponding convergence results for the associated ergodic averages, the investigation of equidistribution properties for Hardy sequences has been carried out several times throughout the literature. Theorem E solves the problem in the abelian case.

Suppose now that we are given a nilmanifold $X = G/\Gamma$, where G is a nilpotent Lie group. We are interested in the behavior of the sequence

$$v(n) = (b_1^{\lfloor a_1(n)\rfloor}\Gamma, \ldots, b_k^{\lfloor a_k(n)\rfloor}\Gamma), \tag{1.5}$$

where $b_1, \ldots, b_k$ are elements of the group G and $a_1, \ldots, a_k$ are Hardy field functions. Notice that this is a sequence on the product nilmanifold X^k.

The most fundamental equidistribution result in the nilpotent case is due to Leibman [39, 40], who showed that if the functions $a_1, \ldots a_k$ are integer polynomials, then we have equidistribution on a "subspace" of X (called a *subnilmanifold*), as long as we restrict the values of n to appropriate arithmetic progressions. This was an important ingredient in his proof that polynomial ergodic averages converge in the case of a single transformation [38]. More specifically, we present the following theorem [39, Theorem B].

Theorem G (Leibman). *Let $X = G/\Gamma$ be a nilmanifold and $x \in X$. Consider the sequence*

$$g(n) = b_1^{q_1(n)} \ldots b_k^{q_k(n)} \tag{1.6}$$

in G, where $b_1, \ldots, b_k \in G$ and $q_1, \ldots, q_k$ are polynomials with integer coefficients. Then, there exists $Q \in \mathbb{N}$, a closed, connected and rational subgroup H of G and points $x_0, \ldots, x_{Q-1} \in X$, such that for every $r \in \{0, \ldots, Q-1\}$ the sequence $g(Qn + r)x$ is equidistributed on the subnilmanifold Hx_r.

Using this theorem, Leibman showed that if $F : X \to \mathbb{C}$ is a continuous function, then the averages

$$\frac{1}{N}\sum_{n=1}^{N} F(g(n)x)$$

converge pointwise for all $x \in X$. In addition, he proved that if G is connected, the equidistribution of the sequence $g(n)\Gamma$ is controlled by the projection of $g(n)\Gamma$ on the "abelianization" $G/[G,G]\Gamma$ of G/Γ, which is a finite-dimensional torus called the horizontal torus of X.

The next breakthrough in this problem[4] was made by Green and Tao in [27], who quantified precisely the behavior of polynomial orbits on nilmanifolds. This new theorem had notable applications in number theory, most notably in establishing the Gowers uniformity of the W-tricked von Mangoldt function. As is evident in Leibman's work, this theorem provides a tight connection between the equidistribution properties of a polynomial sequence on a nilmanifold with its projection to the horizontal torus.

We now return to the setting of more general Hardy sequences. In the case $k = 1$, Frantzikinakis established [12] that if the function $a(t)$ satisfies (P), then the sequence $b^{\lfloor a(n)\rfloor}x$ is equidistributed on the orbit $Y = \overline{\{b^n x \colon n \in \mathbb{N}\}}$ of b for any $b \in G$ and $x \in X$. This implies that the distance of a from rational polynomials is linked to its equidistribution properties, even in the nilpotent case. In the case of several functions of polynomial growth, he also proved that if $a_1, \dots, a_k$ have pairwise distinct growth rates and satisfy

$$\lim_{t\to+\infty}\left|\frac{a_i(t)}{t^{k_i}\log t}\right| = \lim_{t\to+\infty}\left|\frac{t^{k_i+1}}{a_i(t)}\right| = +\infty \tag{1.7}$$

for some $k_i \in \mathbb{N}$, then, for any nilmanifold $X = G/\Gamma$ and $b_1, \dots, b_k \in G$, the sequence

$$\left(b_1^{\lfloor a_1(n)\rfloor}x_1, \dots, b_k^{\lfloor a_k(n)\rfloor}x_k\right)_{n\in\mathbb{N}} \tag{1.8}$$

is equidistributed on $\overline{(b_1^n x_1)}_{n\in\mathbb{N}} \times \cdots \times \overline{(b_k^n x_k)}_{n\in\mathbb{N}}$ for all $x_1, \dots, x_k \in X$. He conjectured that if the linear combinations of the functions $a_1, \dots, a_k$ satisfy the more relaxed assumptions of Theorem 1.1.1, then the sequence in (1.8) is equidistributed on $\overline{(b_1^n x_1)_{n\in\mathbb{N}}} \times \cdots \times \overline{(b_k^n x_k)_{n\in\mathbb{N}}}$. We remark here that all of these theorems do not follow from results for general systems (such as Theorem 1.1.1), because those yield information for the limit in the L^2-sense.

Recently, Richter [49] established another equidistribution theorem for Hardy sequences. Assume, for the sake of exposition, that the underlying Lie group G is connected and simply connected. In this case, there is a natural way to define the elements b^s for $b \in G$ and $s \in \mathbb{R}$. We also denote

$$\nabla - \operatorname{span}\{a_1, \dots, a_k\} = \{c_1 a_1^{(n_1)}(t) + \cdots + c_k a_k^{(n_k)}(t)\colon \ c_i \in \mathbb{R},\ n_i \in \mathbb{N}\cup\{0\}\}.$$

Then, Richter proved that if $a_1, \dots, a_k$ are chosen so that for any function $a \in \nabla - \operatorname{span}\{a_1, \dots, a_k\}$, we have that

$$|a(t) - q(t)| \text{ is bounded or } \lim_{t\to+\infty}\frac{|a(t)-q(t)|}{\log t} = +\infty,$$

for any polynomial $q(t) \in \mathbb{R}[t]$ then, for any commuting elements $b_1, \dots, b_k \in G$, there exists a closed, connected, and rational subgroup H of G and points $x_0, \dots x_{Q-1}$ in X, such that the sequence

$$b_1^{a_1(Qn+r)} \dots b_k^{a_k(Qn+r)}.$$

is equidistributed on the subnilmanifold Hx_r of X for all $r \in \{0, \dots, Q-1\}$.

This theorem is a generalization of Leibman's theorem to the setting of Hardy seqeunces, implying that the orbits are equidistributed on a subspace of X, after we pass to appropriate arithmetic progressions. Using a notion of equidistribution with respect to (weaker) averaging schemes in place

[4]While their theorem was established under the stronger hypothesis that the underlying nilpotent Lie group G is connected and simply connected, one can typically reduce to this case in many applications. We will use this reduction in our arguments as well.

of Cesáro averages[5], Richter weakened the assumptions on $a_1, \dots, a_k$ significantly. Lastly, we remark that while the previous theorem involves the sequences $a_i(n)$ instead of $\lfloor a_i(n) \rfloor$ in the exponents, the statement is in a sense more general since it can be used to prove equidistribution theorems for the sequences $\lfloor a_i(n) \rfloor$.

In order to state our results, we will assume that we have a fixed Hardy field $\mathcal{H}$ that includes the polynomial functions. Removing this last restriction may be possible, though this would certainly complicate our arguments. We will make an exception only for Theorem 1.2.3, since we will have to employ Theorem 1.1.2 that has stronger assumptions on $\mathcal{H}$. We will use the nilmanifolds $\overline{(b^{\mathbb{R}}x)}$ and $\overline{(b^{\mathbb{N}}x)}$ below, which are defined by

$$\overline{(b^{\mathbb{R}}x)} = \overline{\{b^s x \colon s \in \mathbb{R}\}}$$

and similarly for $\overline{(b^{\mathbb{N}}x)}$. We will be able to define $\overline{(b^{\mathbb{R}}x)}$ only under the assumption that the nilpotent Lie group G is simply connected. While their definition as subsets of X is not ambiguous, it is not clear that these sets can be equipped with the structure of a nilmanifold (i.e. a homogeneous space on some nilpotent Lie group). We will prove this assertion in the next chapter and we will keep referring to them as subnilmanifolds in our theorems for now.

Theorem 1.2.1. *[55, Theorem 1.1] Let $\mathcal{H}$ be a Hardy field containing the polynomial functions. Let $a_1, ..., a_k$ be functions in $\mathcal{H}$ that have polynomial growth. Assume that there exists[6] an $\varepsilon > 0$, such that every function $a \in \mathcal{L}(a_1, ..., a_k)$ satisfies*

$$\lim_{t \to +\infty} \frac{|a(t) - q(t)|}{t^{\varepsilon}} = +\infty \quad \text{for any polynomial} \quad q(t) \in \mathbb{Q}[t]. \tag{1.9}$$

Then, we have the following:
(i) For any collection of nilmanifolds $X_i = G_i/\Gamma_i$, elements $b_i \in G_i$ and $x_i \in X_i$, the sequence

$$(b_1^{\lfloor a_1(n) \rfloor} x_1, ..., b_k^{\lfloor a_k(n) \rfloor} x_k)$$

is equidistributed on the nilmanifold $\overline{(b_1^{\mathbb{N}} x_1)} \times \cdots \times \overline{(b_k^{\mathbb{N}} x_k)}$.
(ii) For any collection of nilmanifolds $X_i = G_i/\Gamma_i$ such that the groups G_i are connected, simply connected, elements $b_i \in G_i$ and $x_i \in X_i$, the sequence

$$(b_1^{a_1(n)} x_1, ..., b_k^{a_k(n)} x_k)$$

is equidistributed on the nilmanifold $\overline{(b_1^{\mathbb{R}} x_1)} \times \cdots \times \overline{(b_k^{\mathbb{R}} x_k)}$.

Remark. We will establish the more general statement that if $b_1, \dots, b_k$ commute, the sequence $b_1^{a_1(n)} \cdots b_k^{a_k(n)} \Gamma$ is equidistributed on the nilmanifold $\overline{b_1^{\mathbb{R}} \cdots b_k^{\mathbb{R}} \Gamma}$. The fact that this is indeed a more general statement can be seen by passing to the product nilmanifold $X_1 \times \cdots \times X_k$. A similar assertion holds for Theorem 1.2.2 below and we provide more details on this deduction after Proposition 4.2.1.

In contrast to Theorem E, we have the term t^{ε} in the denominator, which is just out of reach of the conjectured optimal term $\log t$. As an example, using Theorem 1.2.1, we can prove that for any elements $b_1, b_2 \in G$, the sequence $(b_1^{n \log n} \Gamma, b_2^{n^{3/2}} \Gamma)$ is equidistributed on the nilmanifold $(\overline{b_1^{\mathbb{R}} \Gamma}, \overline{b_2^{\mathbb{R}} \Gamma})$, assuming that G satisfies the appropriate connectedness assumptions, since we want these elements to be well defined.

If we have functions that are not linearly independent, then the above theorem fails, as can be seen by noting that the sequence $(n^{3/2}, n^{1/2}, n^{3/2} + n^{1/2})$ is not equidistributed on $\mathbb{T}^3$. However, we can relax the linear independence condition in Theorem 1.2.1 and still obtain a convergence result:

Theorem 1.2.2. *[55, Theorem 1.2] Let $\mathcal{H}$ be a Hardy field containing the polynomial functions. Let $a_1, ..., a_k$ be functions in $\mathcal{H}$ that have polynomial growth. Assume that there exists $\varepsilon > 0$, such that every function $a \in \mathcal{L}(a_1, ..., a_k)$ satisfies either*

$$\lim_{t \to +\infty} \frac{|a(t) - q(t)|}{t^{\varepsilon}} = +\infty \quad \text{for any polynomial} \quad q(t) \in \mathbb{Q}[t], \tag{1.10}$$

[5] A very simple example in the abelian setting is the sequence $\log n$, which is not equidistributed in the standard sense, but is equidistributed with respect to logarithmic averages.

[6] The value of ε depends only on the initial collection $\{a_1, ..., a_k\}$.

or

$$\text{the limit } \lim_{t\to+\infty} a(t) \text{ is a real number.} \tag{1.11}$$

Then, we have the following:
(i) For any collection of nilmanifolds $X_i = G_i/\Gamma_i$*, elements* $b_i \in G_i$*,* $x_i \in X_i$ *and continuous functions* $f_1, ..., f_k$ *with complex values, the averages*

$$\frac{1}{N}\sum_{i=1}^{N} f_1(b_1^{\lfloor a_1(n)\rfloor}x_1)\cdot ... \cdot f_k(b_k^{\lfloor a_k(n)\rfloor}x_k)$$

converge.
(ii) For any collection of nilmanifolds $X_i = G_i/\Gamma_i$ *such that the groups* G_i *are connected, simply connected, elements* $b_i \in G_i$*,* $x_i \in X_i$ *and continuous functions* $f_1, ..., f_k$ *with complex values, the averages*

$$\frac{1}{N}\sum_{i=1}^{N} f_1(b_1^{a_1(n)}x_1)\cdot ... \cdot f_k(b_k^{a_k(n)}x_k)$$

converge.

The main difference between Theorems 1.2.1 and 1.2.2 is that we allow for linear dependencies between the functions $a_1(t), ..., a_k(t)$ in the second case (for example, we may have the functions $(t\log t, t^{3/2}, t^{3/2} + t\log t)$). We will use this theorem and Theorem 1.1.2 to deduce the following.

Theorem 1.2.3. *[55, Theorem 1.3] Let* $\mathcal{H}$ *be a Hardy field that contains the field* $\mathcal{LE}$ *of logarithmico-exponential functions and is closed under composition and compositional inversion of functions (when defined). Furthermore, assume that the functions* $a_1, ..., a_k \in \mathcal{H}$ *are as in Theorem 1.2.2. Then, for any measure preserving system* $(X, \mathcal{X}, \mu, T)$ *and any functions* $f_1, ..., f_k \in L^\infty(\mu)$*, the averages*

$$\frac{1}{N}\sum_{n=1}^{N} T^{\lfloor a_1(n)\rfloor} f_1 \cdot ... \cdot T^{\lfloor a_k(n)\rfloor} f_k \tag{1.12}$$

converge in $L^2(\mu)$*.*

Determining the exact limit of the averages is very difficult in this case and relies on understanding the exact polynomial relations between $a_1, \dots, a_k$, which can be very complicated in the general setting.

Theorems 1.2.1 and 1.2.2 extend the equidistribution result of Frantzikinakis [12], where the functions $a_1, ..., a_k$ were assumed to have different growth rates and satisfy the growth condition in (1.7). On the other hand, our results are complementary to the results in [49], in the sense that each covers collections of functions that are not covered by the other one. The main new cases that are covered in our results (in the case $k \geq 2$) involve functions satisfying a growth condition of the form $t^\ell \prec a(t) \ll t^\ell \log t$, where ℓ is a positive integer. For instance, we can cover all functions of the form $\sum_{i=1}^k c_i t^{a_i}(\log t)^{b_i}$, where $a_i > 0$ and $b_i, c_i \in \mathbb{R}$ (assuming, of course, that the linear combinations of the involved functions satisfy either (1.10) or (1.11)).

As we stated, there are cases covered in the results of [49] that do not follow from the arguments presented here. These examples concern functions that grow slower than fractional powers t^δ, such as the function $(\log t)^A$ for $A > 0$ or the function $\exp(\sqrt{\log t})$. An example that is not covered by Theorem 1.2.2 is the pair of functions $(\log^2 t, t^{3/2})$, which can be covered by the results in [49]. We remark that the conjectured optimal restrictions on the functions $a_1, ..., a_k$ in Theorem 1.2.2 are expected to be that the functions are good for convergence when the system (X, μ, T) is any rotation on some torus $\mathbb{T}^d$. A conjecture of Frantzikinakis appears in [15, Problem 22], although the statement needs to be changed to the following:

Conjecture 1. *Let* $a_1, ..., a_k$ *be functions in* $\mathcal{LE}$ *(or any other Hardy field) with polynomial growth such that for all real numbers* $t_1, ..., t_k \in [0, 1)$*, the averages*

$$\frac{1}{N}\sum_{n=1}^{N} e(t_1\lfloor a_1(n)\rfloor + \cdots + t_k\lfloor a_k(n)\rfloor) \tag{1.13}$$

converge. Then, for any measure preserving system (X,μ,T) *and functions* $f_1,...,f_k\in L^\infty(\mu)$*, the averages*

$$\frac{1}{N}\sum_{n=1}^{N} T^{\lfloor a_1(n)\rfloor} f_1\cdot\ldots\cdot T^{\lfloor a_k(n)\rfloor} f_k \tag{1.14}$$

converge in $L^2(\mu)$ *and, if* (X,μ,T) *is a nilsystem and the functions* $f_1,...,f_k$ *are continuous, then those averages converge pointwise everywhere.*

1.3 Ergodic averages along prime numbers

A general problem in ergodic theory is to prove whether the convergence of the averages (1.1) is still true, if we restrict the range of summation to the primes. More specifically, we want to determine whether the averages

$$\frac{1}{\pi(N)}\sum_{p\in\mathbb{P}:\ p\leq N} T_1^{a_1(p)} f_1\cdot\ldots\cdot T_k^{a_k(p)} f_k \tag{1.15}$$

converge in $L^2(\mu)$ and what is the corresponding limit of these averages. Here, $\pi(N)$ denotes the number of primes less than or equal to N and $\mathbb{P}$ is the set of primes. In this section, we will present results concerning convergence of averages like (1.15) in the case of Hardy sequences and their applications to combinatorics. It will be clear in the results involving prime numbers, that certain arithmetic obstructions (related to polynomial functions) force us to consider the set of shifted primes $\mathbb{P}-1$ (or $\mathbb{P}+1$) in place of $\mathbb{P}$ to prove multiple recurrence results. A simple example is the set $4\mathbb{Z}+2$ which has positive density, but does not contain a pair of the form $(m,m+p)$ with $m\in\mathbb{N}$ and $p\in\mathbb{P}$.

Historically, the first result concerns the case $k=1$ and is due to Sárközy [50]. Using methods from analytic number theory, he showed that sets of positive density contain patterns of the form $(m,m+p-1)$, where p is a prime. Wierdl [57] established the pointwise convergence result of the averages (1.15) in the case $k=1$ and $a_1(n)=n$ and Nair generalized this theorem to polynomials evaluated at primes [46].

In the case of several iterates, Frantzikinakis, Host, and Kra [19] proved that sets of positive density contain 3-term arithmetic progressions whose common difference is a shifted prime. They also showed that the averages in (1.15) converge in the case $k=2$, $T_1=T_2$ and $a_i(n)=in$, $i\in\{1,2\}$. Wooley and Ziegler [59] generalized this to the case of systems with a single transformation and where $a_i(n)$, $i\in\{1,\ldots,k\}$ are polynomials with integer coefficients. Following that, Frantzikinakis, Host, and Kra [20] proved that the Bergelson-Leibman theorem (Theorem D) holds along the shifted primes. In addition, they showed that the averages in (1.15) converge in norm when $a_i(n)$ are integer polynomials, conditional to a conjecture that polynomial ergodic averages converge for several commuting transformations. This last conjecture was subsequently verified by Walsh [56].

Sun obtained convergence and recurrence results in [52] in the case of a single transformation and sequences of the form $i\lfloor an\rfloor$, where $i\in\{1,\ldots,k\}$ with a irrational. Koutsogiannis extended the convergence result of [20] to real polynomials in [33], obtaining recurrence for polynomials with real coefficients rounded to the closest integer. Combinatorial applications along the shifted primes were derived as well through Furstenberg's correspondence principle.

A common theme in all of these results was the reliance on tools built by Green and Tao in their aim to show that primes contain arbitrarily long arithmetic progressions [24]. In particular, a deep theorem known as the Gowers uniformity of the von Mangoldt function (proven by Green and Tao in [25] conditional to conjectures that were subsequently verified in [28] and [26]) is very important in almost all of these results.

Frantzikinakis conjectured that ergodic averages along primes should converge for more general sequences involving fractional powers n^c, such as $\lfloor n^{3/2}\rfloor$, $\left\lfloor n^{\sqrt{e}}\right\rfloor$ as well as for more general Hardy field sequences. To be more precise, he conjectured that the averages

$$\frac{1}{\pi(N)}\sum_{p\in\mathbb{P}:\ p\leq N} T^{\lfloor p^c\rfloor} f_1\cdot\ldots\cdot T^{k\lfloor p^c\rfloor} f_k \tag{1.16}$$

converge in $L^2(\mu)$ for all positive integers k and all positive non-integers c and that the limit is the same as the limit of the averages (1.2). When $0 < c < 1$, the range of $\lfloor p_n^c \rfloor$ contains all sufficiently large integers, so that the multiple recurrence result follows easily. Additionally, the convergence of the previous averages is known in the case $k = 1$ since one can use the spectral theorem and the fact that the sequence $\{p_n^c a\}$ is equidistributed mod 1 for all non-zero $a \in \mathbb{R}$. This last assertion follows from [51] or [58] when $c < 1$ and [41] in the case $c > 1$.

There were significant obstructions to the solution of this problem. One approach would be to modify the comparison method from [20] (concerning polynomials), but the Gowers uniformity of the von Mangoldt function is insufficient to establish this claim. The other approach would be to use the method of characteristic factors, which eventually reduces the task of proving convergence to the case of nilmanifolds. However, this required some equidistribution results on nilmanifolds for the sequence $\lfloor p_n^c \rfloor$, which were very difficult to establish.

A similar conjecture by Frantzikinakis was made for iterates involving distinct fractional powers, such as

$$\frac{1}{\pi(N)} \sum_{p \in \mathbb{P}\colon p \leq N} T^{\lfloor p^{c_1} \rfloor} f_1 \cdot \ldots \cdot T^{\lfloor p^{c_k} \rfloor} f_k$$

for distinct positive non-integer $c_1, \ldots, c_k$. Recently, Frantzikinakis [16] verified that these averages converge in $L^2(\mu)$ to the product of the integrals of the functions $f_1, \ldots, f_k$ in any ergodic system in the more general case where the sequences in the iterates are linearly independent fractional polynomials. These methods relied heavily on the use of the joint ergodicity results in [17] and, thus, the linear independence assumption on the fractional polynomials is necessary. In the same paper, it was conjectured [16, Problem] that the case of fractional polynomials can be generalized to the larger class of Hardy field functions.

We present here our main theorems again under the assumption our Hardy field $\mathcal{H}$ contains the polynomial functions. A few results impose additional assumptions on $\mathcal{H}$ and we state those when necessary. These extra assumptions are a byproduct of convergence results along $\mathbb{N}$.

First of all, we will need to introduce the von Mangoldt function, which will be used in place of the characteristic function of the primes. This is defined by

$$\Lambda(n) = \begin{cases} \log p & , \text{ if } n = p^k \text{ for some prime } p \text{ and } k \in \mathbb{N} \\ 0 & , \text{ otherwise} \end{cases}. \tag{1.17}$$

The function Λ has average 1 by the prime number theorem. Usually, the prime powers with exponents at least 2 contribute a term of significantly lower order in asymptotics, so Λ is morally supported on primes. However, due to the irregularity of the distribution of Λ in residue classes to small moduli, one typically considers a modified version of Λ, called the W-tricked version. To define this, let w be a positive integer and let $W = \prod_{p \leq w, p \in \mathbb{P}} p$. Then, for any integer $1 \leq b \leq W$ with $(b, W) = 1$, we define the W-tricked von Mangoldt function $\Lambda_{w,b}$ by

$$\Lambda_{w,b}(n) = \frac{\phi(W)}{W} \Lambda(Wn + b), \tag{1.18}$$

where ϕ denotes the Euler totient function. This modification can be traced back to [24], where Green and Tao proved that primes contain arbitrarily long arithmetic progressions.

The first theorem will allow us to transfer mean convergence results for Cesàro averages to the prime setting, by establishing a comparison between standard Cesàro averages and averages weighted by the W-tricked von Mangoldt function. The proof requires some equidistribution assumption on the functions a_{ij}, which is characterized through Theorem E.

Theorem 1.3.1. *[36, Theorem 1.1] Let $\mathcal{H}$ be a Hardy field that contains the polynomial functions. Let ℓ, k be positive integers and, for all $1 \leq i \leq k$, $1 \leq j \leq \ell$, let $a_{ij} \in \mathcal{H}$ be functions of polynomial growth such that*

$$\lim_{t \to +\infty} \left| \frac{a_{ij}(t) - q(t)}{\log t} \right| = +\infty \quad \textit{for every polynomial } q(t) \in \mathbb{Q}[t], \tag{1.19}$$

or

$$\lim_{t\to+\infty} |a_{ij}(t) - q(t)| = 0 \quad \text{for some polynomial } q(t) \in \mathbb{Q}[t] + \mathbb{R}. \tag{1.20}$$

Then, for any measure-preserving system $(X, \mathcal{X}, \mu, T_1, \dots, T_k)$ *and functions* $f_1, \dots, f_\ell \in L^\infty(\mu)$*, we have*

$$\lim_{w\to+\infty} \limsup_{N\to+\infty} \max_{\substack{1\le b\le W\\ (b,W)=1}} \Big\| \frac{1}{N} \sum_{n=1}^{N} \big(\Lambda_{w,b}(n) - 1\big) \prod_{j=1}^{\ell} \big(\prod_{i=1}^{k} T_i^{\lfloor a_{ij}(Wn+b) \rfloor}\big) f_j \Big\|_{L^2(\mu)} = 0.$$

Theorem 1.3.1 is used to derive all of our applications on the primes, at least when combined with known results for averages along $\mathbb{N}$. We remark that unlike several of the theorems above or in the literature, there are no linear independence assumptions between the functions a_{ij}, although these assumptions will be necessary in subsequent theorems. The following result, which is effectively a corollary of Theorem 1.3.1, exemplifies how the comparison between averaging schemes works.

Theorem 1.3.2. *[36, Theorem 1.2] Let* $\mathcal{H}$ *be a Hardy field that contains the polynomial functions. Let* ℓ, k *be positive integers,* $(X, \mathcal{X}, \mu, T_1, \dots, T_k)$ *be a measure-preserving system and* $f_1, \dots, f_k \in L^\infty(\mu)$*. Assume that for all* $1 \le i \le k,\ 1 \le j \le \ell$, $a_{ij} \in \mathcal{H}$ *are functions of polynomial growth such that the following conditions are satisfied:*
(a) Each one of the functions $a_{ij}(t)$ *satisfies either* (1.19) *or* (1.20)*.*
(b) For all positive integers W, b*, the averages*

$$\frac{1}{N} \sum_{n=1}^{N} \big(\prod_{i=1}^{k} T_i^{\lfloor a_{i1}(Wn+b) \rfloor}\big) f_1 \cdot \ldots \cdot \big(\prod_{i=1}^{k} T_i^{\lfloor a_{i\ell}(Wn+b) \rfloor}\big) f_\ell \tag{1.21}$$

converge in $L^2(\mu)$*.*

Then, the averages

$$\frac{1}{\pi(N)} \sum_{p\in\mathbb{P}\colon p\le N} \big(\prod_{i=1}^{k} T_i^{\lfloor a_{i1}(p) \rfloor}\big) f_1 \cdot \ldots \cdot \big(\prod_{i=1}^{k} T_i^{\lfloor a_{i\ell}(p) \rfloor}\big) f_\ell \tag{1.22}$$

converge in $L^2(\mu)$*.*

Furthermore, if the averages in (1.21) *converge to the function* $F \in L^\infty(\mu)$ *for all positive integers* W, b*, then the limit in* $L^2(\mu)$ *of the averages* (1.22) *is equal to* F*.*

In our setting, the fact that we require convergence for sequences along arithmetic progressions is typically harmless. Indeed, convergence results along $\mathbb{N}$ typically follow from a growth condition on the implicit functions a_{ij} (such as (1.19)) and it is straightforward to check that the function $a_{ij}(Wt+b)$ satisfies a similar growth condition to $a_{ij}(t)$.

The final part of Theorem 1.3.2 allows us to compute the limit of averages along primes in cases where we have an expression for the limit of the standard Cesàro averages. This is possible, in rough terms, whenever the linear combinations of the functions a_{ij} do not contain polynomials or functions that are approximately equal to a polynomial. The reason for that is that there is no explicit description of the limit of polynomial ergodic averages in a general measure preserving system unless we have some total ergodicity assumptions on the system.

1.3.1 Convergence of ergodic averages along primes

The first application is that the averages in (1.2) converge when $a(n)$ is a Hardy sequence and when we average along primes. The following theorem is a corollary of our comparison and the convergence results of Frantzikinakis [13]. This provides an affirmative answer to [13, Problem 7], which was stated only in the special case of fractional powers $n^c, c \in \mathbb{R}^+ \setminus \mathbb{N}$.

Theorem 1.3.3. *[36, Theorem 1.3] Let* $a \in \mathcal{H}$ *be a function of polynomial growth that satisfies either*

$$\lim_{t\to+\infty} \Big| \frac{a(t) - cq(t)}{\log t} \Big| = +\infty \quad \text{for every } c \in \mathbb{R} \text{ and every } q \in \mathbb{Z}[t], \tag{1.23}$$

or

$$\lim_{t\to+\infty} |a(t) - cq(t)| = d \text{ for some } c, d \in \mathbb{R} \text{ and some } q \in \mathbb{Z}[t]. \tag{1.24}$$

Then, for any positive integer k, any measure-preserving system $(X, \mathcal{X}, \mu, T)$ and functions $f_1, \ldots, f_k \in L^\infty(\mu)$, we have that the averages

$$\frac{1}{\pi(N)} \sum_{p\in\mathbb{P}:\, p\leq N} T^{\lfloor a(p)\rfloor} f_1 \cdot \ldots \cdot T^{k\lfloor a(p)\rfloor} f_k \tag{1.25}$$

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

In particular, if a satisfies (1.23), the limit of the averages in (1.25) is equal to the limit in $L^2(\mu)$ of the averages

$$\frac{1}{N}\sum_{n=1}^{N} T^n f_1 \cdot \ldots \cdot T^{kn} f_k.$$

Remark. The reader may observe that condition iii) in Theorem F is the only one missing above. However, the proof of Theorem 1.3.1 becomes much more complicated if we allow this case as well.

The following theorem concerns the "jointly ergodic" case for one transformation, which refers to the setting when we have convergence to the product of the integrals in ergodic systems. Observe that the assumptions on the Hardy field functions are exactly the same as in Theorem 1.1.1 and the conclusion is the same apart from the fact that we average along the prime numbers. This theorem generalizes the theorem of Frantzikinakis [16, Theorem 1.1] and gives a positive answer to [16, Problem].

Theorem 1.3.4. *[36, Theorem 1.4] Let $\mathcal{H}$ be a Hardy field that contains $\mathcal{LE}$ and is closed under composition and compositional inversion of functions, when defined. For a positive integer k, let $a_1, \ldots, a_k$ be functions of polynomial growth and assume that every non-trivial linear combination a of them satisfies*

$$\lim_{t\to+\infty} \left|\frac{a(t) - q(t)}{\log t}\right| = +\infty \text{ for every } q(t) \in \mathbb{Z}[t]. \tag{1.26}$$

Then, for any measure-preserving system $(X, \mathcal{X}, \mu, T)$ and functions $f_1, \ldots, f_k \in L^\infty(\mu)$, we have that

$$\lim_{N\to+\infty} \frac{1}{\pi(N)} \sum_{p\in\mathbb{P}:\, p\leq N} T^{\lfloor a_1(p)\rfloor} f_1 \cdot \ldots \cdot T^{\lfloor a_k(p)\rfloor} f_k = \tilde{f}_1 \cdot \ldots \cdot \tilde{f}_k, \tag{1.27}$$

where $\tilde{f}_i := \mathbb{E}(f_i|\mathcal{I}(T)) = \lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} T^n f_i$ and the convergence is in $L^2(\mu)$.

We remark that we can also transfer the convergence result of Theorem 1.2.3 to primes.

In the case of several commuting transformations, results on the limiting behavior for averages for Hardy sequences along $\mathbb{N}$ are few. The only known convergence result is due to Frantzikinakis [14, Theorem 2.3], which we now transfer to the prime setting. By a shift-invariant Hardy field, we are referring to a Hardy field such that $a(t+h) \in \mathcal{H}$ for any $h \in \mathbb{Z}$ and function $a(t) \in \mathcal{H}$.

Theorem 1.3.5. *[36, Theorem 1.5] Let $k \in \mathbb{N}$, $\mathcal{H}$ be a shift-invariant Hardy field that contains the polynomial functions, $a_1, \ldots, a_k$ be functions in $\mathcal{H}$ with pairwise distinct growth rates and such that there exist integers $d_i \geq 0$ satisfying*

$$\lim_{t\to+\infty} \left|\frac{a_i(t)}{t^{d_i}\log t}\right| = \lim_{t\to+\infty} \left|\frac{t^{d_i+1}}{a_i(t)}\right| = +\infty.$$

Then, for any system $(X, \mathcal{X}, \mu, T_1, \ldots, T_k)$ and functions $f_1, \ldots, f_k \in L^\infty(\mu)$, we have

$$\lim_{N\to+\infty} \frac{1}{\pi(N)} \sum_{p\in\mathbb{P}:\, p\leq N} T_1^{\lfloor a_1(p)\rfloor} f_1 \cdot \ldots \cdot T_k^{\lfloor a_k(p)\rfloor} f_k = \tilde{f}_1 \cdot \ldots \cdot \tilde{f}_k,$$

where $\tilde{f}_i := \mathbb{E}(f_i|\mathcal{I}(T_i)) = \lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} T_i^n f_i$ and the convergence is in $L^2(\mu)$.

While there are more restrictions compared to Theorem 1.3.4, we note that Theorem 1.3.5 covers at least the case of distinct fractional powers, i.e. when $a_i(t) = t^{c_i}$ for distinct, positive non-integers c_i.

1.3.2 Applications to multiple recurrence and combinatorics

In this subsection, we will translate the previous convergence results to multiple recurrence results and then combine them with Furstenberg's correspondence principle (Theorem B) to derive combinatorial applications. Due to arithmetic obstructions arising from polynomials, we have to work with the set of shifted primes in some cases. In addition, it was observed in [33] that in the case of real polynomials, one needs to work with the rounding to the closest integer function instead of the floor function. Indeed, even in the case of sequences of the form $\lfloor ap(n)+b \rfloor$, explicit conditions that describe multiple recurrence are very complicated (cf. [13, Footnote 4]). We will denote by $[[x]]$ the closest integer to x.

Our first application relates to the averages appearing in Theorem 1.3.3.

Theorem 1.3.6. *[36, Theorem 1.6] Let $a \in \mathcal{H}$ be a function of polynomial growth. Then, for any measure-preserving system $(X, \mathcal{X}, \mu, T)$, $k \in \mathbb{N}$, and set A with positive measure we have the following:*
(a) If a satisfies (1.23), we have

$$\lim_{N\to+\infty} \frac{1}{\pi(N)} \sum_{p\in\mathbb{P}:\, p\leq N} \mu(A \cap T^{-\lfloor a(p)\rfloor}A \cap \cdots \cap T^{-k\lfloor a(p)\rfloor}A) > 0.$$

(b) If a satisfies (1.24) with $cp(0)+d=0$,[7] then for any set A with positive measure, the set

$$\Big\{n \in \mathbb{N}:\ \mu\big(A \cap T^{-[[a(n)]]}A \cap \cdots \cap T^{-k[[a(n)]]}A\big) > 0\Big\}$$

has non-empty intersection with the sets $\mathbb{P}-1$ or $\mathbb{P}+1$.

This theorem and Furstenberg's correspondence principle yield the following corollary.

Corollary 1.3.7. *[36, Corollary 1.7] For any set $E \subseteq \mathbb{N}$ of positive upper density, $k \in \mathbb{N}$, and function $a \in \mathcal{H}$ of polynomial growth, the following holds:*
(a) If a satisfies (1.23), we have

$$\liminf_{N\to+\infty} \frac{1}{\pi(N)} \sum_{p\in\mathbb{P}:\, p\leq N} \bar{d}\big(E \cap (E - \lfloor a(p)\rfloor) \cap \cdots \cap (E - k\lfloor a(p)\rfloor)\big) > 0.$$

(b) If a satisfies (1.24) with $cp(0)+d=0$, then the set

$$\big\{n \in \mathbb{N}:\ \bar{d}\big(E \cap (E - [[a(n)]]) \cap \cdots \cap (E - k[[a(n)]])\big) > 0\big\}$$

has non-empty intersection with the sets $\mathbb{P}-1$ or $\mathbb{P}+1$.

Specializing to the case where $a(n)=n^c$ where c is a positive non-integer, Theorem 1.3.3 and part (a) of Theorem 1.3.6 provide an affirmative answer to [15, Problem 27].

We remark here that in part (a) of both Theorem 1.3.6 and Corollary 1.3.7 one can evaluate the function a along the affine shifts $ap+b$ for $a,b \in \mathbb{Q}$ with $a \neq 0$. This follows from the fact that the function $a_i(at+b)$ satisfies (1.23) as well, if a_i does. However, the shifts $p-1$ and $p+1$ are the only correct ones in part (b) of Theorem 1.3.6 as there are simple counterexamples otherwise.

Now, we state the recurrence result obtained by Theorem 1.3.4.

Theorem 1.3.8. *[36, Theorem 1.8] Let $k \in \mathbb{N}$, $\mathcal{H}$ be a Hardy field that contains $\mathcal{LE}$ and is closed under composition and compositional inversion of functions, when defined, and suppose $a_1,\dots,a_k \in \mathcal{H}$ are functions of polynomial growth whose non-trivial linear combinations satisfy (1.3). Then, for any measure-preserving system $(X, \mathcal{X}, \mu, T)$, and set A with positive measure, we have that*

$$\lim_{N\to+\infty} \frac{1}{\pi(N)} \sum_{p\in\mathbb{P}:\, p\leq N} \mu\big(A \cap T^{-\lfloor a_1(p)\rfloor}A \cap \cdots \cap T^{-\lfloor a_k(p)\rfloor}A\big) \geq \big(\mu(A)\big)^{k+1}.$$

[7] We have to use an assumption that the polynomial has no constant term, in order to obtain a recurrence result.

Corollary 1.3.9. *[36, Corollary 1.9] For any $k \in \mathbb{N}$, set $E \subseteq \mathbb{N}$ of positive upper density, Hardy field $\mathcal{H}$ and functions $a_1, \ldots, a_k \in \mathcal{H}$ as in Theorem 1.3.8, we have*

$$\liminf_{N\to+\infty} \frac{1}{\pi(N)} \sum_{p\in\mathbb{P}\colon p\le N} \bar{d}\big(E \cap (E - \lfloor a_1(p)\rfloor) \cap \cdots \cap (E - \lfloor a_k(p)\rfloor)\big) \ge \big(\bar{d}(E)\big)^{k+1}.$$

In particular, we conclude that for any set $E \subseteq \mathbb{N}$ with positive upper density and $a_1, \ldots, a_k$ as above, the set

$$\{n \in \mathbb{N}\colon \text{ there exists } m \in \mathbb{N} \text{ such that } m, m + \lfloor a_1(n)\rfloor, \ldots, m + \lfloor a_k(n)\rfloor \in E\}$$

has non-empty intersection with the set $\mathbb{P}$.

The following is a multidimensional analog of Theorem 1.3.8 and relies on the convergence result of Theorem 1.3.5.

Theorem 1.3.10. *[36, Theorem 1.10] Let $k \in \mathbb{N}$, $\mathcal{H}$ be a shift-invariant Hardy field and suppose that $a_1, \ldots, a_k \in \mathcal{H}$ are functions of polynomial growth that satisfy the hypotheses of Theorem 1.3.5. Then, for any system $(X, \mathcal{X}, \mu, T_1, \ldots, T_k)$ and set A with positive measure, we have that*

$$\lim_{N\to+\infty} \frac{1}{\pi(N)} \sum_{p\in\mathbb{P}\colon p\le N} \mu\big(A \cap T_1^{-\lfloor a_1(p)\rfloor} A \cap \cdots \cap T_k^{-\lfloor a_k(p)\rfloor} A\big) \ge \big(\mu(A)\big)^{k+1}.$$

Lastly, we present the corresponding combinatorial application of our last multiple recurrence result. Given a set $E \subseteq \mathbb{Z}^d$, its *upper density* is given by

$$\bar{d}(E) := \limsup_{N\to+\infty} \frac{|E \cap \{-N, \ldots, N\}^d|}{(2N+1)^d}.$$

Corollary 1.3.11. *[36, Corollary 1.11] For any $k \in \mathbb{N}$, set $E \subseteq \mathbb{Z}^d$ of positive upper density, Hardy field $\mathcal{H}$ and functions $a_1, \ldots, a_k \in \mathcal{H}$ as in Theorem 1.3.10 and vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{Z}^d$, we have*

$$\liminf_{N\to+\infty} \frac{1}{\pi(N)} \sum_{p\in\mathbb{P}\colon p\le N} \bar{d}\big(E \cap (E - \lfloor a_1(p)\rfloor \mathbf{v}_1) \cap \cdots \cap (E - \lfloor a_k(p)\rfloor \mathbf{v}_k)\big) \ge \big(\bar{d}(E)\big)^{k+1}.$$

Equidistribution in nilmanifolds

In this part, we present some results relating to pointwise convergence in nilmanifolds along Hardy sequences evaluated at primes. We have the following theorem that translates results from the setting of averages along $\mathbb{N}$ to primes, similar to Theorem 1.3.2.

Theorem 1.3.12. *[36, Theorem 1.12] Let k be a positive integer. Assume that $a_1, \ldots, a_k \in \mathcal{H}$ are functions of polynomial growth, such that the following conditions are satisfied:*
(a) For every $1 \le i \le k$, the function $a_i(t)$ satisfies either (1.19) or (1.20).
(b) For all positive integers W, b, any nilmanifold $Y = H/\Delta$, pairwise commuting elements $u_1, \ldots, u_k$ and points $y_1, \ldots, y_k \in Y$, the sequence

$$\Big(u_1^{\lfloor a_1(Wn+b)\rfloor} y_1, \ldots, u_k^{\lfloor a_k(Wn+b)\rfloor} y_k\Big)$$

is equidistributed on the nilmanifold $\overline{(u_1^{\mathbb{Z}} y_1)} \times \cdots \times \overline{(u_k^{\mathbb{Z}} y_k)}$.

Then, for any nilmanifold $X = G/\Gamma$, pairwise commuting elements $g_1, \ldots, g_k \in G$ and points $x_1, \ldots, x_k \in X$, the sequence

$$\Big(g_1^{\lfloor a_1(p_n)\rfloor} x_1, \ldots, g_k^{\lfloor a_k(p_n)\rfloor} x_k\Big)_{n\in\mathbb{N}},$$

where p_n denotes the n-th prime, is equidistributed on the nilmanifold $\overline{(g_1^{\mathbb{Z}} x_1)} \times \cdots \times \overline{(g_k^{\mathbb{Z}} x_k)}$.

Instead of the "pointwise convergence" assumption (b), one can replace it with a weaker convergence (i.e. in the L^2-sense) hypothesis. However, we will not benefit from this in applications, so we opt to not state our results in that setup.

In the case of a polynomial function, a convergence result along primes follows by combining [26, Theorem 7.1] (which is the case of linear polynomials) and the fact that any polynomial orbit on a nilmanifold can be lifted to a linear orbit of a unipotent affine transformation on a larger nilmanifold (an argument due to Leibman [39]). Nonetheless, in this case, we do not have a nice description for the orbit of this polynomial sequence.

On the other hand, this theorem is the first result on equidistribution in higher-step nilmanifolds (along primes) for sequences such as $\lfloor p_n^c \rfloor$, with $c > 1$ a non-integer (this was previously unknown even in the simplest case of one fractional power).

All of the pointwise convergence theorems that we mentioned above can be transferred to the prime setting. As an application, we have the following corollary of Theorem 1.3.12 and Theorem 1.2.1. The term invariant under affine shifts refers to a Hardy field $\mathcal{H}$ for which $a(Wt+b) \in \mathcal{H}$ whenever $a \in \mathcal{H}$, for all $W, b \in \mathbb{N}$.

Corollary 1.3.13. *[36, Corollary 1.13] Let k be a positive integer, $\mathcal{H}$ be a Hardy field invariant under affine shifts, and suppose that $a_1, \dots, a_k \in \mathcal{H}$ are functions of polynomial growth, for which there exists an $\varepsilon > 0$, so that every non-trivial linear combination a of them satisfies*

$$\lim_{t \to +\infty} \left| \frac{a(t) - q(t)}{t^{\varepsilon}} \right| = +\infty \text{ for every } q(t) \in \mathbb{Z}[t]. \tag{1.28}$$

Then, for any collection of nilmanifolds $X_i = G_i/\Gamma_i$ $i = 1, \dots, k$, elements $g_i \in G_i$ and points $x_i \in X_i$, the sequence

$$\big(g_1^{\lfloor a_1(p_n) \rfloor} x_1, \dots, g_k^{\lfloor a_k(p_n) \rfloor} x_k\big)_{n \in \mathbb{N}},$$

where p_n denotes the n-th prime, is equidistributed on the nilmanifold $\overline{(g_1^{\mathbb{Z}} x_1)} \times \cdots \times \overline{(g_k^{\mathbb{Z}} x_k)}$.

The condition on $\mathcal{H}$ is necessary, because we want to apply Theorem 1.2.1 for sequences of the form $a(Wn+b)$ for all choices of $W, b \in \mathbb{N}$. The assumption on $\mathcal{H}$ can, in principle, be dropped since the arguments in our proofs rely on some growth assumptions on the functions a_i which also hold for their shifted versions, but this would complicate the statement and proof of Theorem 1.2.1.

Our corollary implies that the sequence

$$\big(g_1^{\lfloor p_n^{c_1} \rfloor} x_1, \dots, g_k^{\lfloor p_n^{c_k} \rfloor} x_k\big)$$

is equidistributed on the subnilmanifold $\overline{(g_1^{\mathbb{Z}} x_1)} \times \cdots \times \overline{(g_k^{\mathbb{Z}} x_k)}$ of $X_1 \times \cdots \times X_k$, for any distinct positive non-integers $c_1, \dots, c_k$ and for all points $x_i \in X_i$. This is stronger than the result of Frantzikinakis [16] that establishes convergence in the L^2-sense (for linearly independent fractional polynomials). This result is novel even in the simplest case $k = 1$. Furthermore, we remark that in the case $k = 1$ we can actually replace (1.28) with the optimal condition that $a(t) - q(t)$ grows faster than $\log t$, for all $q(t)$ that are real multiples of integer polynomials, using the results from [12].

Ideas and organization of the proofs

In general, we have a very strong and developed theory in determining convergence of averages like (1.1), which is the Host-Kra structure theory. This works really well in the case of polynomial sequences, because we can use elementary tools such as the Cauchy-Schwarz and van der Corput inequalities to reduce the complexity of the polynomials appearing in the iterates. In very broad terms, this allows to replace the sequence $q(n)$ (where q is a polynomial) with sequences of the form $q(n+h) - q(n)$ where h is a parameter over which we average. The significant gain of this procedure is that this new polynomial has smaller degree than the previous one. Thus, after finitely many steps, this process will terminate and we will be able to bound ergodic averages involving polynomials by an appropriate Host-Kra seminorm (defined in the next chapter), which, on its own, is sufficient to reduce the problem to the case of nilmanifolds. In short, there is some sort of induction happening in the case of polynomials that reduces their degrees after finitely many iterations of this procedure.

In the case of Hardy sequences, our main task is to transfer our initial problem to a setting involving only polynomial iterates. This can be achieved through Taylor expansion in an appropriate range. More specifically, the best way to achieve this is to show that if we average over a short interval of the form $[N, N + L(N)]$ where $L(t)$ is a positive function that grows to infinity slower than linearly, then we can approximate our Hardy sequence $a(n)$ by its Taylor polynomial in the range $[N, N + L(N)]$. Of course, this polynomial varies with N, but the methods of the polynomial case can still be applied here. If we can bound averages over a narrow range like $[N, N + L(N)]$, then a very simple argument shows that this bound holds for averages in the long ranges such as $[1, N]$. In some cases, we may need to consider more complicated averaging schemes (i.e considering a double average over the parameter N as well) but the main principle underpinning these arguments is still the same.

There are several differences regarding the proofs of the results in the three previous sections. The results in Section 1.1 rely on bounding ergodic averages involving Hardy sequences by Host-Kra seminorms, which is equivalent to proving Theorem 1.1.2. This is done by repeatedly using the van der Corput inequality and an induction scheme originally due to Bergelson [1] (called PET induction in the literature). The remaining theorems of this section follow from these bounds and the joint ergodicity criterion of Frantzikinakis. The joint ergodicity criterion has two conditions that need to be verified. The first necessary condition is that our averages are bounded by a Host-Kra seminorm (which will occupy the bulk of the proof) and an equidistribution (mod 1) assumption on the sequences involved, which will follow easily from Theorem E.

The results involving convergence on nilmanifolds will be attacked by a reduction to a problem involving polynomial sequences through Taylor expansion. Since these polynomials vary with the underlying short interval, we will need a quantitative equidistribution criterion for polynomial orbits on a nilmanifold, which is already known due to Green and Tao [27]. However, this theorem requires several technical definitions relating to nilmanifolds, so we have to postpone its statement till the next chapter.

In order to prove the results along primes, we will use similar approximations as in the case of averages along $\mathbb{N}$. First of all, we will use some very recent number theoretic input from [44] that establishes the Gowers uniformity of the von Mangoldt function in short intervals. However, there are still several complications arising when studuying averages weighted by a von Mangoldt weight, since the von Mangoldt function is unbounded. This creates some problems when trying to completely eliminate the error term of the Taylor polynomials in the iterates so that we can have genuine polynomial sequences. This is circumevented through a series of equidistribution arguments, which at the end will allow us to reduce our problem to averages with iterates of the form $\lfloor p_N(n) \rfloor$, where p_N are polynomials with real coefficients. If the integer parts were not present, then the argument of Frantzikinakis, Host, and Kra ([19] and [20]) that handles the case of integer polynomials could be applied. In order to achieve this, we will pass to an extension of the system $(X, \mathcal{X}, \mu, T_1, \ldots, T_k)$, wherein the actions T_i are lifted to $\mathbb{R}$-actions (also called measure-preserving flows) and the integer parts are removed. This argument was used by Koutsogiannis in order to tackle the case of polynomials with real coefficients and has its origins in [6] and [42].

There are a lot of technical details missing that cannot possibly be fitted in this short discussion. In particular, the whole inductive procedure that we have to follow below is very complicated and notationally heavy. For this reason, we have several examples in Chapters 3, 4 and 5, which all handle very simple cases of the theorems discussed in this chapter. Chapter 3 will concern the proods of all the results of Section 1.1, Chapter 4 will contain the proofs of Section 1.2 and, lastly, we will include the proofs of the remaining results of Section 1.3 in Chapter 5.

1.3.3 Notational conventions

Throughout this thesis, we denote with $\mathbb{N} = \{1, 2, \ldots\}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{P}$ the sets of natural, integer, rational, real, complex numbers and prime numbers respectively. We denote the one dimensional torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, the exponential phases $e(t) = e^{2\pi i t}$, while $\|x\|_{\mathbb{T}} = d(x, \mathbb{Z})$, $[[x]]$, $\lfloor x \rfloor$, $\lceil x \rceil$, and $\{x\}$ are the distance of x from the nearest integer, the nearest integer to x, the greatest integer which is less or equal to x, the smallest integer which is greater or equal to x, and the fractional part of x respectively.

We let $\mathbf{1}_A$ denote the characteristic function of a set A and $|A|$ is its cardinality. For any integers a, Q we use the symbol $a\ (Q)$ to denote the residue class a modulo Q. Therefore, the notation $\mathbf{1}_{a\ (Q)}$ refers to the characteristic function of the set of those integers, whose residue when divided by Q is equal to a.

For two sequences a_n, b_n, we say that b_n *dominates* a_n and write $a_n \prec b_n$ or $a_n = o(b_n)$, when a_n/b_n goes to 0, as $n \to +\infty$. In addition, we write $a_n \ll b_n$ or $a_n = O(b_n)$, if there exists a positive constant C such that $|a_n| \leq C|b_n|$ for large enough n. When we want to denote the dependence of the constant C on some parameters $h_1, \dots, h_k$, we will use the notation $a_n = O_{h_1,\dots,h_k}(b_n)$. In the case that $b_n \ll a_n \ll b_n$, we shall write $a_n \sim b_n$. We say that a_n and b_n have the same growth rate when the limit of $\frac{a_n}{b_n}$, as $n \to +\infty$ exists and is a non-zero real number. We use a similar notation and terminology for asymptotic relations when comparing functions of a real variable t.

Under the same setup as in the previous paragraph, we say that the sequence a_n *strongly dominates* the sequence b_n if there exists $\delta > 0$ such that

$$\frac{a_n}{b_n} \gg n^{\delta}.$$

In this case, we write $b_N \lll a_N$, or $a_N \ggg b_N$.[8] We use similar terminology and notation for functions on a real variable t.

Finally, for any sequence $(a(n))$, we will use the averaging notation

$$\mathop{\mathbb{E}}_{n\in S} a(n) = \frac{1}{|S|}\sum_{n\in S} a(n)$$

to denote averages over a finite non-empty set S. We will typically work with averages over the integers in a specified interval, whose endpoints may not necesseraly be integers. We will use the symbol $\mathcal{E}_\mu(f|\mathcal{Y})$ to denote the conditional expectation of a function f with respect to the σ-algebra $\mathcal{Y}$.

We will use the letters b, g, u, w to denote elements of a Lie group and we will use either bold letters to indicate vector-valued quantities.

Notation involving cubes

Given a positive integer s, we will denote by $[[s]]$ the set $\{0,1\}^s$ of ordered s-tuples of zeroes and ones, which contains 2^s elements (which we refer to as cubes). For elements of cube sets $[[s]]$ only, we will use the notation $\underline{\varepsilon}$ instead of bold letters. For convenience, we will write $\underline{0}, \underline{1}$ for the elements $(0,0,...0)$ and $(1,1,...,1)$ of $[[s]]$ respectively. We will also define $|\underline{\varepsilon}|$ to be the sum of elements of $\underline{\varepsilon}$. For a finite set Y, we will similarly use the notation $Y^{[[s]]}$ to denote the set Y^{2^s}. Each element $\mathbf{h} \in Y^{[[s]]}$ can be represented as $\mathbf{h} = (h_{\underline{\varepsilon}}, \underline{\varepsilon} \in [[s]])$ where each $h_{\underline{\varepsilon}}$ belongs to Y. For complex numbers z, we define the operator $\mathcal{C}^k z$, where $\mathcal{C}^k z := z$, if k is an even number and $\mathcal{C}^k z := \bar{z}$ otherwise.

[8] This notation is non-standard, so we may refer back to this part quite often throughout the text.

Chapter 2

Background

2.1 Preliminaries on Hardy fields

In this section, we describe some basic properties of Hardy fields. The main advantage when working with functions in a Hardy field (instead of just the C^∞ functions) is that any two functions $f, g \in \mathcal{H}$ are comparable. That means that the limit

$$\lim_{t\to\infty} \frac{f(t)}{g(t)}$$

exists (possibly $\pm\infty$) and thus it makes sense to talk about and compare their growth rates. The fact that the limit exists follows from the fact that $\mathcal{H}$ is a field and the fact that every function in $\mathcal{H}$ has a limit. In addition, since every function in our Hardy field has a multiplicative inverse, we can easily infer that it is eventually monotone and, therefore, has constant sign eventually.

It will be crucial in the proof of the theorems of Section 1.1 to assume that $\mathcal{H}$ is closed under composition and compositional inversion of functions, when defined. More precisely, if $f, g \in \mathcal{H}$ are such that $\lim\limits_{t\to+\infty} g(t) = +\infty$, then we have that $f \circ g \in \mathcal{H}$ and $g^{-1} \in \mathcal{H}$. The Hardy field $\mathcal{LE}$ does not have this property. This can be achieved by working with the Hardy field $\mathcal{P}$ of Pfaffian functions [32], which contains $\mathcal{LE}$ and satisfies the previously mentioned assumptions. This field can be defined inductively as follows:
i) Let $\mathcal{P}_1$ be the set of the smooth functions satisfying the differential equation $f' = p(t, f)$ for some polynomial p with integer coefficients.
ii) Let $\mathcal{P}_k$ be the set of the smooth functions satisfying the differential equation $f' = p(t, f_1, ..., f_k)$ for some polynomial p with integer coefficients and $f_i \in P_i$ for $1 \le i \le k-1$. Then $\mathcal{P}$ contains all germs at infinity of the set $\cup_{i=1}^{\infty} \mathcal{P}_i$.

From now on, we will assume that $\mathcal{H}$ has all the above properties. In the appendix, we have gathered some lemmas regarding growth rates of functions in $\mathcal{H}$, which will play a crucial role in the approximations in the following sections.

Finally, we give some definitions for functions whose growth rate is of particular interest.

Definition 2.1.1. *We say that a function $a \in \mathcal{H}$ has* polynomial growth *if there exists a positive integer d such that $a(t) \ll t^d$. The minimal value of d that can be chosen to satisfy this inequality will be called the* degree *of a. We say that a has* sub-linear growth rate *(or is* sub-linear*), if $a(t) \prec t$. We say that a function $a \in \mathcal{H}$ has* sub-fractional growth rate *(or is* sub-fractional*), if for all $\delta > 0$, we have $f(t) \ll t^\delta$.*

Functions that are sub-fractional behave differently from super-fractional functions in terms of the Taylor expansion. That is if we have an interval of the from $[N, N + N^c]$ where $c < 1$, then a simple calculation using the mean value theorem shows that we have that $|a(n) - a(N)| = o_N(1)$ for every $n \in [N, N + N^c]$. Thus, these functions are essentially constant in short intervals of length approximately N^c. This will create the need to separate these functions from super-fractional in our iterates.

Some simple examples of sub-linear functions are $\sqrt{t}$, $e^{\sqrt{\log t}}$ and $\log^3(t)$. Among these, the functions $e^{\sqrt{\log t}}$ and $\log^3(t)$ are also sub-fractional, while the first one is not sub-fractional.

Definition 2.1.2. *We will call a function $f \in \mathcal{H}$ of polynomial growth* strongly non-polynomial*, if there exists a non-negative integer d, such that*

$$t^d \prec f(t) \prec t^{d+1}.$$

For example, the functions $t^{3/2}$ and $\log^3(t)$ are strongly non-polynomial, while the function $t^2+\sqrt{t}$ is not.

2.1.1 Lemmas on growth rates of Hardy sequences

Let us fix a Hardy field $\mathcal{H}$ that contains the polynomials. Firstly, we will need a basic lemma that relates the growth rate of a Hardy field function of polynomial growth with the growth rate of its derivative. To do this, we recall a lemma due to Frantzikinakis [12, Lemma 2.1], as well as [54, Proposition A.1].

Lemma 2.1.3. *Let $a \in \mathcal{H}$ satisfy $t^{-m} \prec a(t) \prec t^m$ for some positive integer m and assume that $a(t)$ does not converge to a non-zero constant as $t \to +\infty$. Then,*

$$\frac{a(t)}{t(\log t)^2} \prec a'(t) \ll \frac{a(t)}{t}.$$

Sketch of proof. Firstly, observe that since $a(t)$ goes to either 0 or $\pm\infty$, L' Hospital's rule implies that

$$\lim_{t\to\infty} \frac{\log|a(t)|}{\log t} = \lim_{t\to+\infty} \frac{ta'(t)}{a(t)}.$$

The limit on the right-hand side exists because $a(t), a'(t)$ and t all belong to $\mathcal{H}$. Since $t^{-m} \ll |a(t)| \ll t^m$, we conclude that the limit on the left and side is finite. This implies that

$$\frac{ta'(t)}{a(t)} \ll 1 \implies a'(t) \ll \frac{a(t)}{t}.$$

To prove the remaining inequality, it suffices to show that the limit of $\frac{t(\log t)^2 a'(t)}{a(t)}$ as $t \to +\infty$ is infinite. If that is not the case, then we would have

$$(\log|a(t)|)' \ll \frac{1}{t(\log t)^2}.$$

Integrating this, we deduce that

$$\log|a(t)| \ll \frac{1}{\log t} + c$$

for some $c \in \mathbb{R}$, which implies that $\log|a(t)|$ is bounded, which contradicts the assumption that $a(t)$ does not converge to a finite limit. □

Observe that if a function $a(t)$ satisfies the growth inequalities in the hypothesis of this lemma, then the function $a'(t)$ satisfies $\frac{t^{-1-m}}{\log^2 t} \prec a'(t) \prec t^{m-1}$. Therefore, we deduce the relations $t^{-m-2} \prec a'(t) \prec t^{m+2}$, which implies that the function $a'(t)$ satisfies a similar growth condition. Provided that the function $a'(t)$ does not converge to a non-zero constant as $t \to +\infty$, the above lemma can then be applied to the function $a'(t)$. In particular, we can show that if a function has polynomial growth, then after sufficiently many differentiations, we will arrive at a function that converges to zero, as $t \to +\infty$.

When a function $a(t)$ is strongly non-polynomial and dominates the logarithmic function $\log t$, one can get a nice ordering relation for the growth rates of consecutive derivatives. This is the content of the following proposition.

Proposition 2.1.4. *Let $a \in \mathcal{H}$ be a function of polynomial growth that is strongly non-polynomial and also satisfies $a(t) \succ \log t$. Then, for all sufficiently large $k \in \mathbb{N}$, we have*

$$1 \prec \left|a^{(k)}(t)\right|^{-\frac{1}{k}} \prec \left|a^{(k+1)}(t)\right|^{-\frac{1}{k+1}} \prec t.$$

Remark. The proof of Proposition 2.1.4 establishes the fact that if a satisfies the previous hypotheses, then the derivatives of a do not converge to a non-zero constant and, thus, always satisfy the conditions of Lemma 2.1.3.

Proof. The function a has non-vanishing derivatives of all orders, since it is not a polynomial. Let d be an integer, such that $t^d \prec a(t) \prec t^{d+1}$. Then, Lemma 2.1.3 implies that $|a^{d+1}(t)| \to 0$. Therefore, for any $k \geq d+1$, we have $a^{(k)}(t) \prec 1$. This gives the leftmost part of the required inequality. In particular, $(d+1)$ is minimal among the integers k, for which $f^{(k)}(t)$ converges to 0.

To prove the rightmost inequality of the proposition, it is sufficient to prove that

$$a^{(d+1)}(t) \succ t^{-d-1}.$$

For $k \geq d+1$, the result then follows by successive applications of L' Hospital's rule. In the case $d=0$, the above relation also follows easily from L'Hospital's rule. Therefore, we may assume that $d \geq 1$. Now, since a is strongly non-polynomial, we have that the function $a^{(d)}(t)$ goes to infinity. We will show that

$$g'(t) \gg \frac{g(t)}{t \log^2 t} \tag{2.1}$$

where g is any one of the functions $a, a', ..., a^{(d)}$. The result then will follow by noting that

$$a^{(d+1)(t)} \gg \frac{a(t)}{t^{d+1}(\log t)^{2d+2}} \gg \frac{1}{t(\log t)^{2d+2}} \succ \frac{1}{t^{d+1}}.$$

Equation (2.1) follows by applying Lemma 2.1.3 and noting that we have $|g(t)| \to +\infty$ (namely, g does not converge to a non-zero constant), since the original function a dominates the function t^d.

It remains to establish the middle part, namely that if $k \geq d+1$, then

$$|a^{(k+1)}(t)|^k \prec |a^{(k)}(t)|^{k+1}.$$

However, we have $|a^{(k+1)}(t)|^k \ll |a^{(k)}(t)|^k/t^k$ by Lemma 2.1.3 and we easily get the conclusion by combining this relation with the relation $t^{-k} \prec a^{(k)}(t)$ that we established in the previous step. □

This proposition is the first step we use to show that a strongly non-polynomial function $a(t)$ can be approximated by polynomials in short intervals. Indeed, assume that a positive sub-linear function $L(t)$ satisfies

$$\left|a^{(k)}(t)\right|^{-\frac{1}{k}} \prec L(t) \prec \left|a^{(k+1)}(t)\right|^{-\frac{1}{k+1}} \tag{2.2}$$

for some sufficiently large $k \in \mathbb{N}$ (large enough so that the inequalities in Proposition 2.1.4 hold). In particular, this implies that $\lim\limits_{t\to+\infty} a^{(k+1)}(t) = 0$ and the convergence is monotone, since $a^{(k+1)}(t)$ is eventually monotone.

Using the Taylor expansion around the point N, we can write

$$a(N+h) = a(N) + ha'(N) + \cdots + \frac{h^k a^{(k)}(N)}{k!} + \frac{h^{k+1} a^{(k+1)}(\xi_{N,h})}{(k+1)!} \text{ for some } \xi_{N,h} \in [N, N+h]$$

for every $0 \leq h \leq L(N)$. However, we observe that

$$\left|\frac{h^{k+1} a^{(k+1)}(\xi_{N,h})}{(k+1)!}\right| \leq \frac{L(N)^{k+1}|a^{(k+1)}(N)|}{(k+1)!} = o_N(1),$$

where we used the fact that $|a^{(k+1)}(t)| \to 0$ monotonically. Therefore, we have

$$a(N+h) = a(N) + ha'(N) + \cdots + \frac{h^k a^{(k)}(N)}{k!} + o_N(1),$$

which implies that the function $a(N+h)$ is essentially a polynomial in h.

The final lemma implies that if the function $L(t)$ satisfies certain growth assumptions, then a strongly non-polynomial function $a(t)$ will be approximated by a polynomial of some degree k. Namely, we can always find $k \in \mathbb{N}$ so that the inequalities (2.2) are satisfied.

Proposition 2.1.5. *Let $a \in \mathcal{H}$ be a strongly non-polynomial function of polynomial growth, such that $a(t) \succ \log t$. Assume that $L(t)$ is a positive sub-linear function, such that $1 \prec L(t) \ll t^{1-\varepsilon}$ for some $\varepsilon > 0$. Then, there exists a non-negative integer k depending on the function $a(t)$ and $L(t)$, such that*

$$\left|a^{(k)}(t)\right|^{-\frac{1}{k}} \prec L(t) \prec \left|a^{(k+1)}(t)\right|^{-\frac{1}{k+1}},$$

where we adopt the convention that $\left|a^{(k)}(t)\right|^{-\frac{1}{k}}$ denotes the constant function 1, when $k = 0$.

Proof. We split the proof into two cases depending on whether a is sub-fractional or not.

Assume first that $a(t) \ll t^{\delta}$ for all $\delta > 0$. We will establish the claim for $k = 0$. This means that functions that are sub-fractional become essentially constant when restricted to intervals of the form $[N, N + L(N)]$. The left inequality is obvious. Furthermore, since $a(t) \prec t^{\varepsilon}$, Lemma 2.1.3 implies that

$$a'(t) \prec \frac{1}{t^{1-\varepsilon}} \ll \frac{1}{L(t)},$$

which yields the desired result.

Assume now that $a(t) \succ t^{\delta}$ for some $\delta > 0$. Observe that, in this case, we have that

$$\left|a^{(k)}(t)\right|^{-\frac{1}{k}} \prec \left|a^{(k+1)}(t)\right|^{-\frac{1}{k+1}}$$

for k large enough, due to Proposition 2.1.4. We also consider the integer d, such that $t^d \prec a(t) \prec t^{d+1}$. This number exists because the function a is strongly non-polynomial.

If $L(t) \prec \left|a^{(d+1)}(t)\right|^{-\frac{1}{d+1}}$, then the claim holds for $k = d$, since $\left|a^{(d)}(t)\right|^{-\frac{1}{d}} \prec 1 \prec L(t)$.

It suffices to show that there exists $k \in \mathbb{N}$, such that $L(t) \prec \left|a^{(k+1)}(t)\right|^{-\frac{1}{k+1}}$, which, in turn, follows if we show that

$$t^{1-\varepsilon} \prec \left|a^{(k+1)}(t)\right|^{-\frac{1}{k+1}} \tag{2.3}$$

for some $k \in \mathbb{N}$. We can rewrite the above inequality as $a^{(k+1)}(t) \prec t^{(k+1)(\varepsilon-1)}$. However, since the function $a(t)$ is strongly non-polynomial and $a(t) \succ \log t$, the functions $a^{(k)}(t)$ satisfy the hypotheses of Lemma 2.1.3 (see also Remark 2.1.1). Therefore, iterating the aforementioned lemma, we deduce that

$$a^{(k+1)}(t) \ll \frac{a(t)}{t^{k+1}}.$$

Hence, it suffices to find k such that $a(t) \ll t^{(k+1)\varepsilon}$ and such a number exists, because the function $a(t)$ has polynomial growth. □

Remark. The condition $L(t) \prec t^{1-\varepsilon}$ is necessary. For example, if $a(t) = t \log t$ and $L(t) = \frac{t}{\log t}$, then for any $k \in \mathbb{N}$, we can write

$$(N + h)\log(N + h) = N \log N + \cdots + \frac{C_1 h^k}{N^{k-1}} + \frac{C_2 h^{k+1}}{\xi_{N,h}^{k}}$$

for every $0 \leq h \leq \frac{N}{\log N}$ and some numbers $C_1, C_2 \in \mathbb{R}$. However, there is no positive integer k for which the last term in this expansion can be made to be negligible since $\frac{N}{\log N} \succ N^{\frac{k}{k+1}}$ for all $k \in \mathbb{N}$. Essentially, in order to approximate the function $t \log t$ in these specific short intervals, one would be forced to use the entire Taylor series instead of some appropriate cutoff.

In the proofs of our results, we will need a version of Proposition 2.1.5 for several functions simultaneously. However, we will need a different version of this simultaneous Taylor approximation depending on whether we work with averages along $\mathbb{N}$ or with averages along primes. Therefore, we state and prove these propositions in their corresponding chapters.

2.2 Background in ergodic theory

2.2.1 Systems, ergodicity and factors

A *measure preserving system is a probability space* $(X, \mathcal{X}, \mu)$ *equipped with an invertible measure preserving transformation* T. *We call a system* ergodic, if the only T-invariant functions in $L^\infty(\mu)$ are the constant ones. The system $(X, \mathcal{X}, \mu, T)$ is called *weak-mixing*, if the product system $(X \times X, \mathcal{X} \times \mathcal{X}, \mu \times \mu, T \times T)$ is ergodic.

More generally, let G be a group. A *measure-preserving G-action* on a Lebesgue probability space $(X, \mathcal{X}, \mu)$ is an action on X by measure-preserving maps T_g for every $g \in G$ such that, for all $g_1, g_2 \in G$, we have $T_{g_1 g_2} = T_{g_1} \circ T_{g_2}$. In this thesis, we will only need to consider actions by the additive groups of $\mathbb{Z}$ or $\mathbb{R}$. Throughout the following sections, we will also refer to $\mathbb{R}$-actions as *measure-preserving flows*. In the case of $\mathbb{Z}$-actions, we follow the usual notation and write T^n to indicate the map T_n.

We say the system $(Y, \mathcal{Y}, \nu, S)$ is a *factor* of $(X, \mathcal{X}, \mu, T)$ (or that $(X, \mathcal{X}, \mu, T)$ is an *extension* of $(Y, \mathcal{Y}, \nu, S)$), if there exist $X' \subset X$, $Y' \subset Y$ of full measure that are invariant under T and S respectively and a map $p : X' \to Y'$ such that $\nu = \mu \circ p^{-1}$ and $p \circ T(x) = S \circ p(x)$ for all $x \in X'$. If p is a bijection, we say that the two systems are *isomorphic*. A factor of the system $(X, \mathcal{X}, \mu, T)$ corresponds to a T-invariant sub-σ-algebra of $\mathcal{X}$ (in the above example this σ-algebra is $p^{-1}(\mathcal{Y})$). Therefore, one can simply think of factors as T-invariant sub σ-algebras. From now on, we will often omit the σ-algebra $\mathcal{X}$ from the quadruple $(X, \mathcal{X}, \mu, T)$ when there is no confusion.

The σ-algebra spanned by T-invariant sets will be called the *invariant algebra or the* invariant factor of the system and will be denoted by $\mathcal{I}(T)$. It is the smallest σ-algebra making the T-invariant functions measurable. A system is ergodic if and only if the invariant factor is trivial.

An *eigenfunction* of the system (X, μ, T) is a function satisfying the relation $Tf = \lambda f$ almost everywhere, where λ is a complex number. The σ-algebra spanned by eigenfunctions is called the *Kronecker factor* of the system. One can show that a system is weak-mixing if and only if the Kronecker factor is trivial.

2.2.2 Host-Kra seminorms, structure factors and Gowers norms

Let (X, μ, T) be an invertible measure preserving system and let $f \in L^\infty(\mu)$. We define the *Host-Kra uniformity seminorms* inductively as follows:

$$\|\!|f|\!\|_{0,T} := \int f \, d\mu$$

and, for $s \in \mathbb{Z}^+$,

$$\|\!|f|\!\|_{s+1,T}^{2^{s+1}} := \lim_{H \to \infty} \mathop{\mathbb{E}}_{0 \le h \le H} \|\!|\bar{f} \cdot T^h f|\!\|_{s,T}^{2^s}. \tag{2.4}$$

When there is no confusion, we will omit the transformation from the subscripts.

The existence of the limits above was proven in [29] in the ergodic case (for the non-ergodic case, see [30] for a proof) and it was also established that the $\|\!|\cdot|\!\|_s$ are indeed seminorms for $s \neq 0$. The seminorms are increasing, which means that for any bounded function f we have $\|\!|f|\!\|_s \le \|\!|f|\!\|_{s+1}$ for all $s \ge 0$. In the case $s = 1$, we only have an easier description of the seminorm, namely that $\|\!|f|\!\|_1 = \|\mathcal{E}_\mu(f|\mathcal{I}(T))\|_{L^2(\mu)}$. Furthermore, it is easy to prove that $\|\!|\bar{f} \otimes f|\!\|_{s,T\times T} \le \|\!|f|\!\|_{s+1,T}^2$, where $\bar{f} \otimes f$ denotes the function $(x, y) \to \overline{f(x)} f(y)$ on $(X \times X, \mu \times \mu, T \times T)$.

Expanding the inductive definition above, we infer that the seminorms $\|\!|f|\!\|_s$ take the form

$$\|\!|f|\!\|_s^{2^s} = \lim_{H_s \to +\infty} \mathop{\mathbb{E}}_{0 \le h_s \le H_s} \cdots \lim_{H_1 \to +\infty} \mathop{\mathbb{E}}_{0 \le h_1 \le H_1} \int \prod_{\underline{\varepsilon} \in [[s]]} \mathcal{C}^{|\underline{\varepsilon}|} T^{h_1 \varepsilon_1 + \cdots + h_s \varepsilon_s} f \, d\mu \tag{2.5}$$

where we use the notation $\underline{\varepsilon} = (\varepsilon_1 \ldots, \varepsilon_k)$ for every $\underline{\varepsilon} \in [[k]]$. We refer the reader to our notational conventions for the symbols regarding cubes. We see that there is some sort of cubic structure in these seminorms. For instance, we have that

$$\|\!|f|\!\|_2^4 = \lim_{H_2 \to +\infty} \mathop{\mathbb{E}}_{0 \le h_2 \le H_2} \lim_{H_1 \to +\infty} \mathop{\mathbb{E}}_{0 \le h_1 \le H_1} \int f \cdot \overline{T^{h_1} f} \cdot \overline{T^{h_2} f} \cdot T^{h_1 + h_2} f \, d\mu.$$

Host-Kra proved that the limits over the parameters $h_1, \ldots, h_s$ in (2.5) can be taken to be simultaneous instead of iterated (for a proof in the ergodic case, see [29]).

The importance of these seminorms lies in the following two properties. Firstly, it was shown by Host and Kra in the same article that, for all $s \geq 1$, the seminorms $|||f|||_{s,T}$ define a factor $Z_{s-1}(X)$ of X, which is characterized by the following property:

$$f \perp L^2(Z_{s-1}(X)) \iff |||f|||_{s,T} = 0.$$

The factors $Z_s(X)$ form an increasing sequence of factors, which follows from the monotonicity property of the seminorms. The factor $Z_0(X)$ corresponds to the invariant factor of the system, while the factor $Z_1(X)$ is the Kronecker factor of the system, when the system is ergodic. Finally, in the case of weak-mixing systems, it can be shown that all the factors $Z_s(X)$ are trivial.

The second important property of these Host-Kra seminorms hinges on the structure of the $Z_s(X)$ factors, which is contained in the next celebrated theorem of Host and Kra.

Theorem G (Host-Kra[29]). *Let (X, μ, T) be an ergodic system. Then, the factor $Z_s(X)$ is an inverse limit of s-step nilsystems.*

The last property implies that there exists an increasing sequence of T-invariant sub-σ-algebras $Z_s(n), n \in \mathbb{N}$ that span Z_s, such that the factor $Z_s(n)$ is isomorphic as a system to an s-step nilsystem (we give the exact definition of an nilystem in the next section).

The main strategy in proving convergence results (in the single iterate case) is as follows: we bound the L^2-norm of our averages by the Host-Kra seminorms of all functions involved using the Cauchy-Schwarz and van der Corput inequalities. This implies that if one of the functions $f_1, \ldots, f_k$ in (1.1) is orthogonal to $Z_s(X)$ (for the value of s that the inductive procedure above provides), then our averages are zero. Thus, if we write each of the functions as the sum of its projections to $L^2(Z_s(X))$ and its orthogonal complement, then a telescoping argument implies that we can replace $f_1, \ldots, f_k$ by its projections to $L^2(Z_s(X))$. In this case, we say that $Z_s(X)$ is characteristic for the averages in question. An example of a theorem of this form is 1.1.2. After this step has been completed then the structure theorem and a simple approximation argument allow us to reduce our problem to the case that our system is a nilsystem. Then, one has to prove convergence in this system exploiting the algebraic structure of the nilsystem.

The Host-Kra seminorms are related to the Gowers norms, introduced by Gowers in his proof of Szemerédi's theorem [23]. These are defined typically for sequences along groups and along the integers with slight modifications. We will mostly need the latter definition in our proofs.

Let N be a positive integer and let $f : \mathbb{Z}_N \to \mathbb{C}$ be a function. For any positive integer s, we define the *Gowers uniformity norm* $\|f\|_{U^s(\mathbb{Z}_N)}$ inductively by

$$\|f\|_{U^1(\mathbb{Z}_N)} = \Big|\underset{n\in\mathbb{Z}_N}{\mathbb{E}} f(n)\Big|$$

and for $s \geq 2$,

$$\|f\|^{2^s}_{U^s(\mathbb{Z}_N)} = \underset{h\in\mathbb{Z}_N}{\mathbb{E}} \|\overline{f(\cdot)}f(\cdot + h)\|^{2^{s-1}}_{U^{s-1}(\mathbb{Z}_N)}.$$

A straightforward computation implies that

$$\|f\|_{U^s(\mathbb{Z}_N)} = \Big(\underset{\underline{h}\in\mathbb{Z}_N^s}{\mathbb{E}}\ \underset{n\in\mathbb{Z}_N}{\mathbb{E}} \prod_{\underline{\varepsilon}\in\{0,1\}^s} \mathcal{C}^{|\underline{\varepsilon}|} f(n + \underline{h}\cdot\underline{\varepsilon})\Big)^{\frac{1}{2^s}}.$$

Observe that these seminorms are very similar to the Host-Kra seminorms (our system is $\mathbb{Z}_N$ with the shift map $Tx = x + 1 \pmod N$).

It can be shown that, for all $s \geq 2$, $\|\cdot\|_{U^s(\mathbb{Z}_N)}$ is a norm and that

$$\|f\|_{U^s(\mathbb{Z}_N)} \leq \|f\|_{U^{s+1}(\mathbb{Z}_N)}$$

for any function f on $\mathbb{Z}_N$ [30, Chapter 6].

In the setting of sequences defined along the integers, we can define the Gowers norms analogously. For any $s \geq 1$ and a finitely supported sequence $f(n), n \in \mathbb{Z}$, we define the *unnormalized Gowers uniformity norm*

$$\|f\|_{U^s(\mathbb{Z})} = \Big(\sum_{\underline{h}\in\mathbb{Z}^s}\sum_{n\in\mathbb{Z}}\prod_{\underline{\varepsilon}\in\{0,1\}^s}\mathcal{C}^{|\underline{\varepsilon}|}f(n+\underline{h}\cdot\underline{\varepsilon})\Big)^{\frac{1}{2^s}} \tag{2.6}$$

and for a bounded interval $I \subset \mathbb{R}$, we define

$$\|f\|_{U^s(I)} = \frac{\|f\cdot \mathbf{1}_I\|_{U^s(\mathbb{Z})}}{\|\mathbf{1}_I\|_{U^s(\mathbb{Z})}}. \tag{2.7}$$

First of all, observe that a simple change of variables in the summation in (2.7) implies that for $X \in \mathbb{Z}$

$$\|f\|_{U^s(X,X+H]} = \|f(\cdot + X)\|_{U^s[1,H]}.$$

Evidently, we want to compare uniformity norms on the interval $[1, H]$ with the corresponding norms on the abelian group $\mathbb{Z}_H$. To this end, we will use the following lemma, whose proof can be found in [30, Chapter 22, Proposition 11].

Lemma 2.2.1. *Let s be a positive integer and $N, N' \in \mathbb{N}$ with $N' \geq 2N$. Then, for any sequence $\big(f(n)\big)_{n\in\mathbb{Z}}$, we have*

$$\|f\|_{U^s[1,N]} = \frac{\|f\cdot 1_{[1,N]}\|_{U^s(\mathbb{Z}_{N'})}}{\|1_{[1,N]}\|_{U^s(\mathbb{Z}_{N'})}}.$$

We will need a final lemma that implies that the Gowers uniformity norm is smaller when the sequence is evaluated along arithmetic progressions.

Lemma 2.2.2. *Let $u(n)$ be a sequence of complex numbers. Then, for any integer $s \geq 2$ and any positive integers $0 \leq a \leq Q-1$, we have*

$$\|u(n)\mathbf{1}_{a\ (Q)}(n)\|_{U^s(X,X+H]} \leq \|u(n)\|_{U^s(X,X+H]},$$

for all integers $X \geq 0$ and all $H \geq 1$.

Proof. We set $u_X(n) = u(X+n)$, so that we can rewrite the norm on the left-hand side as $\|u_X(n)\mathbf{1}_{a\ (Q)}(X+n)\|_{U^s[1,H]}$. Observe that the function $\mathbf{1}_{a\ (Q)}(n)$ is periodic modulo Q. Thus, treating it as a function in $\mathbb{Z}_Q$, we have the Fourier expansion

$$\mathbf{1}_{a\ (Q)}(n) = \sum_{\xi\in\mathbb{Z}_Q}\widehat{\mathbf{1}}_{a\ (Q)}(\xi)e\Big(\frac{n\xi}{Q}\Big),$$

for every $0 \leq n \leq Q-1$, and this can be extended to hold for all $n \in \mathbb{Z}$ due to periodicity. Furthermore, we have the bound

$$|\widehat{\mathbf{1}}_{a\ (Q)}(\xi)| = \frac{1}{Q}\Big|e\Big(\frac{a\xi}{Q}\Big)\Big| \leq \frac{1}{Q}.$$

Applying the triangle inequality, we deduce that

$$\|u_X(n)\mathbf{1}_{a\ (Q)}(X+n)\|_{U^s[1,H]} \leq \sum_{\xi\in\mathbb{Z}_Q}|\widehat{\mathbf{1}}_{a\ (Q)}(\xi)|\cdot\Big\|u_X(n)e\Big(\frac{(X+n)\xi}{Q}\Big)\Big\|_{U^s[1,H]}.$$

However, it is immediate from (2.6) that the U^s-norm is invariant under multiplication by linear phases, for every $s \geq 2$. Therefore, we conclude that

$$\|u_X(n)\mathbf{1}_{a\ (Q)}(X+n)\|_{U^s[1,H]} \leq \|u_X(n)\|_{U^s[1,H]} = \|u(n)\|_{U^s(X,X+H]},$$

which is the desired result. □

The primary utility of the Gowers uniformity norms is the fact that they arise naturally in complexity reduction arguments that involve multiple ergodic averages with polynomial iterates, just like the Host-Kra seminorms. We will use them in problems that involve polynomial ergodic averages weighted by a sequence $(a(n))_{n\in\mathbb{N}}$ in order to bound the averages by the Gowers norm of the weight $a(n)$. The sequence $a(n)$ will be the modified von Mangoldt function in our applications.

2.2.3 Joint ergodicity of sequences

Let $a_1(n), ..., a_k(n)$ be sequences of integers. Following the terminology in [17], we call these sequences *jointly ergodic*, if for any ergodic measure preserving system (X, μ, T) and functions $f_1, ..., f_k \in L^\infty(\mu)$, we have

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} T^{a_1(n)}f_1 \cdot ... \cdot T^{a_k(n)}f_k = \int f_1 \, d\mu \cdot .... \cdot \int f_k \, d\mu,$$

where convergence takes place in $L^2(\mu)$. We also give the following definitions:

Definition 2.2.3. *We say that a collection of sequences $a_1, ..., a_k$ of integers:*
i) is good for seminorm estimates*, if for every ergodic system (X, μ, T) there exists an $s \in \mathbb{N}$, such that if $f_1, ..., f_k \in L^\infty(\mu)$ and $\|f_\ell\|_s = 0$ for some $\ell \in \{1, ..., k\}$, then*[1]

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} T^{a_1(n)}f_1 \cdot ... \cdot T^{a_k(n)}f_k = 0$$

in $L^2(\mu)$.
ii) is good for equidistribution*, if for all $t_1, ..., t_k \in [0, 1)$, not all of them zero, we have*

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} e(t_1 a_1(n) + \cdots + t_k a_k(n)) = 0.$$

The main result of Frantzikinakis in [17], which we are also going to use is the following:

Theorem H. *[17, Theorem 1.1] Let $a_1, ..., a_k$ be a collection of sequences of integers. Then, the following are equivalent:*
i) The sequences $a_1, ..., a_k$ are jointly ergodic.
ii) The sequences $a_1, ..., a_k$ are good for seminorm estimates and good for equidistribution.

2.3 Background on nilmanifolds

2.3.1 Definitions and basic properties

Here, we present the basic definitions and tools concerning nilmanifolds. We follow the notation and symbols used in [30] and most of the theorems can be found in this book. The reader that is interested in the general theory of nilpotent Lie groups can also consult [9].

Let G be a topological group. A subgroup H of a topological group G is called *discrete*, if there is a cover of H by open sets of G, such that each of these open sets contains exactly one element of H. It is called *co-compact* if the quotient topology makes G/H a compact space. We call a subgroup with both of the above properties uniform and we will use the letters Γ or Δ to denote such subgroups.

Let G be a k-step nilpotent Lie group and Γ be a uniform subgroup. The space $X = G/\Gamma$ is called a k-step nilmanifold.

Let b be any element in G. Then, b acts on G by left multiplication. Let m_X be the image of the Haar measure of G on X under the natural projection map. Then, m_X is invariant under the action of the element b (and therefore the action of G). If we set $T(g\Gamma) = (bg)\Gamma$, then the transformation T is called a *nilrotation*, and (X, m_X, T) is called a *nilsystem*. If the transformation T is ergodic, we say that b acts ergodically on the nilmanifold X. It can be proven that b acts ergodically on X if and only the sequence $(b^n x)_{n\in\mathbb{N}}$ is dense on X for all $x \in X$ (see, for instance, [30, Chapter 11]).

A simple example of a nilsystem is a rotation on $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, that is the space $\mathbb{T}$ with the Lebesgue measure and the map $x \to x + a(mod 1)$, where $a \in [0, 1]$. An example in the non-abelian case is the Heisenberg system. This is the space

$$\begin{pmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix} \Big/ \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}$$

[1] In [17], this property is called "*very good for seminorm estimates*".

where the multiplication is the usual matrix multiplication. This is easily seen to be a 2-step nilmanifold. Given any matrix element on the original Lie group we can find an element on the fundamental domain through the map

$$\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & \{x\} & \{z - x\lfloor y \rfloor\} \\ 0 & 1 & \{y\} \\ 0 & 0 & 1 \end{pmatrix}.$$

In this case, it can be shown that a rotation by a matrix element $\begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix}$ is ergodic if and only if $1, a$ and b are rationally independent. Furthermore, the Kronecker factor of the system in this case is the rotation on $\mathbb{T}^2$ by $(x, y) \to (x + a, y + b)$ where addition is done modulo 1. Observe that the condition that a rotation on the Heisenberg manifold is ergodic is equivalent to the condition that the induced rotation on the Kronecker factor is ergodic. We will see that this holds for general nilmanifolds.

Let x_n be a sequence of elements on $X = G/\Gamma$. We say that x_n is *equidistributed* on $X = G/\Gamma$ if and only if for every continuous function $F : X \to \mathbb{C}$, we have

$$\lim_{N\to+\infty} \mathop{\mathbb{E}}_{1\le n\le N} F(x_n) = \int F dm_X$$

where m_X is the (normalized) Haar measure of X. We say that x_n is well-distributed on $X = G/\Gamma$ if and only if for every continuous function $F : X \to \mathbb{C}$, we have

$$\lim_{N\to+\infty} \sup_{M\in\mathbb{N}} \mathop{\mathbb{E}}_{M\le n\le M+N} F(x_n) = \int F dm_X.$$

A rational subgroup H is a subgroup of G such that $H \cdot e_X$ is a closed subset of $X = G/\Gamma$, where e_X is the identity element of X. Equivalently, $H\Gamma$ is a closed subset of the space G. This, also, implies that H must be closed in G (see [30, Chapter 10, Lemma 14]).

A *subnilmanifold* of $X = G/\Gamma$ is a set of the form Hx, where H is a closed subgroup of the Lie group G, $x \in X$ and such that Hx is closed in X. Observe that if H is a rational subgroup of G, then He_X is a subnilmanifold of X.

Horizontal torus and characters

Assume $X = G/\Gamma$ is a k-step nilmanifold with G connected and simply connected and consider the subgroup $G_2 = [G, G]$. The nilmanifold $Z = G/(G_2\Gamma)$ is called the horizontal torus of X. We observe that Z is a connected, compact Abelian Lie group, and thus isomorphic to some torus $\mathbb{T}^d$. For a $b \in G$, it can be shown that the nilrotation induced by b is ergodic, if and only if the induced action of b on Z is ergodic. More precisely, we have the following theorem due to Leibman (see the theorem in section 2.17 of [39]). Given an amenable group A and a homomorphism $\phi : A \to G$, then A acts on $X = G/\Gamma$ by translations $(\phi(u))(x) = \phi(u)x$. This action is called ergodic if the only functions invariant under the action of A are constant.

Theorem I (Leibman)**.** *Let $X = G/\Gamma$ be a connected nilmanifold, A be an amenable group acting on X and let G° be the connected component of the identity element. Then, the action of A is ergodic if and only if the induced action on the factor torus $[G^\circ, G^\circ]\backslash X$ is ergodic.*

This theorem was proven by Parry [47] in the case of a connected group G. The nilmanifold $[G^\circ, G^\circ]\backslash X$ is referred to as the *maximal factor torus* of X. It can be shown that the induced action of an element b on the maximal factor torus is isomorphic to a unipotent affine transformation on a finite-dimensional torus. This, however, complicates things when G is not connected. However, notice that when G is connected, then the maximal factor torus is the same as the horizontal torus. Thus, we have the following corollary.

Corollary 2.3.1. *Let $X = G/\Gamma$ be a nilmanifold with G connected. Then, the action of b is ergodic if and only if the induced action on the horizontal torus is ergodic.*

A horizontal character χ is a continuous group morphism $\chi: G \to \mathbb{C}$, such that $\chi(g\gamma) = \chi(g)$ for all $\gamma \in \Gamma$. We observe that χ also annihilates G_2 and therefore descends to the horizontal torus Z. Thus, under the natural projection map π, χ becomes a character on some torus $\mathbb{T}^d$. We will often use the notation $\chi \circ \pi$ when working in the horizontal torus, while we reserve the letter χ to denote the same character in the original group G.

2.3.2 Nilorbits and Ratner's theorem

Let G be a connected, simply-connected nilpotent Lie group. It is well known that the exponential map exp from the Lie algebra of G to G is a diffeomorphism (see, for instance, [9, Theorem 1.2.1]). In particular, it is a bijection between G and its Lie algebra $\mathfrak{g}$. For $b \in G$ and $t \in \mathbb{R}$ we can then define the element b^t as the unique element of G satisfying $b^t = \exp(tL)$, where $L \in \mathfrak{g}$ satisfies $\exp(L) = b$. As a corollary of Ratner's theorem [48], we get the following:

Lemma 2.3.2. *Let $X = G/\Gamma$ be a nilmanifold with G connected and simply connected. For any elements $b_1, ..., b_k \in \Gamma$, we have that the set*

$$\overline{b_1^{\mathbb{R}} \cdots b_k^{\mathbb{R}} \Gamma} = \overline{\{b_1^{t_1} \cdots b_k^{t_k} \Gamma \colon t_1, ..., t_k \in \mathbb{R}\}}$$

is a subnilmanifold of $X = G/\Gamma$ with a representation H/Δ, for some closed, connected and rational subgroup H of G that contains the elements $b_1^s, ..., b_k^s$ for all $s \in \mathbb{R}$ and Δ is a uniform subgroup of H.

We call the set $\overline{\{b^t\Gamma \colon t \in \mathbb{R}\}}$ the nil-orbit of the element b. We will analogously denote by $\overline{b^{\mathbb{Z}}\Gamma}$ the set $\overline{\{b^n\Gamma \colon n \in \mathbb{Z}\}}$ and $\overline{b^{\mathbb{N}}\Gamma} = \overline{\{b^n\Gamma \colon n \in \mathbb{N}\}}$.

We establish the following lemma, which will be necessary for our proofs.

Lemma 2.3.3. *Let $X = G/\Gamma$ be a nilmanifold and let $b_1, ..., b_k \in \Gamma$ be any pairwise commuting elements. Then, there exists a real number t such that*

$$\overline{b_1^{\mathbb{R}} \dots b_k^{\mathbb{R}} \Gamma} = \overline{\{b_1^{n_1 t} \dots b_k^{n_k t} \Gamma \colon n_1, ..., n_k \in \mathbb{Z}\}}.$$

Proof. We want to find some $t \in \mathbb{R}$ so that the sequence

$$\phi_t(n_1, ..., n_k) = b_1^{n_1 t} \dots b_k^{n_k t}$$

is equidistributed on the nilmanifold $Y = \overline{b_1^{\mathbb{R}} \dots b_k^{\mathbb{R}} \Gamma}$. By Lemma 2.3.2, Y has a representation as H/Δ, where H is connected, simply connected and rational. Observe that ϕ_t naturally induces a $\mathbb{Z}^k$ action on Y by $(\phi_t(n_1, ..., n_k), h\Delta) \to b_1^{n_1 t} \dots b_k^{n_k t} h\Delta$. It is sufficient to show that this $\mathbb{Z}^k$-action is ergodic on Y, since this implies that $Y = \overline{\{\phi_t(\mathbf{n})y, \mathbf{n} \in \mathbb{Z}^k\}}$ for all $y \in Y$. However, using Theorem I, the above action is ergodic if and only if it is ergodic on the horizontal torus Z of Y, which is homeomorphic to some torus $\mathbb{T}^d$. Equivalently, if we denote by $(b_{i,1}, ..., b_{i,d})$ the projection of the point $b_i\Gamma$ on Z, then we need to check whether the sequence

$$\big(t(n_1 b_{i,1} + \cdots + n_k b_{k,1}), ..., t(n_1 b_{1,d} + \cdots + n_k b_{k,d})\big)$$

is dense on $\mathbb{T}^d$. It suffices to choose t so that $1/t$ is rationally independent of any integer combination of the coordinates $b_{i,j}$. This completes the proof. □

2.3.3 Polynomial sequences on nilmanifolds

In the proofs of our theorems, we will approximate our Hardy sequences with polynomials. We define polynomial sequences here, though we work in a more general setting by defining polynomial sequences through filtrations. We will need this to state the quantitative results on the equidistribution of polynomial orbits in larger generality.

Definition 2.3.4. *A filtration $G_\bullet$ of degree d on a nilpotent Lie group G is a sequence of closed connected subgroups*

$$G = G^{(0)} = G^{(1)} \supseteq G^{(2)} \supseteq \cdots \supseteq G^{(d)} \supseteq G^{(d+1)} = e_G,$$

such that $[G^{(i)}, G^{(j)}] \subseteq G^{(i+j)}$ for all $i, j \geq 0$. The filtration is called rational if all groups $G^{(i)}$ appearing in the above sequence are rational subgroups of G. A polynomial sequence on G with respect to the above filtration is a sequence $g(n)$ such that, for all positive integers $h_1, ..., h_k$, we have that the sequence $\partial_{h_1} \dots \partial_{h_k} g$ takes values in $G^{(k)}$, for all $k \in \mathbb{N}$, where ∂_h denotes the "differencing operator" that maps the sequence $(g(n))_{n\in\mathbb{N}}$ to the sequence $(g(n+h)(g(n))^{-1})_{n\in\mathbb{N}}$.

An example of a filtration is the lower central series of the group G. For our purposes, we will only need to consider polynomial sequences of the form

$$v(n) = b_1^{p_1(n)} \cdot \ldots \cdot b_k^{p_k(n)} \tag{2.8}$$

where $b_i \in G$ for all $1 \leq i \leq k$ and p_i are real polynomials. Note that the terms $b_i^{p_i(n)}$ are well defined, due to our connectedness assumptions. To see that this is indeed a polynomial sequence with our initial definition, we construct a specific filtration on G. We assume that G is k-step nilpotent and we also denote the maximum degree among the polynomials p_i as d. We consider the filtration (of degree dk) $G_\bullet = (G^{(i)})_{0\leq i\leq dk}$, where $G^{(i)} = G_{\lfloor i/d \rfloor +1}$ and G_j are the commutator subgroups of G. This is a rational filtration because all commutator subgroups of G are rational (see [30, Chapter 10, Proposition 22] for the proof). Then, the sequence $v(n)$ in (2.8) is a polynomial sequence with respect to this filtration. We direct the reader to the discussion after [27, Corollary 6.8], where these last observations were made originally. We will also call the projected sequence $v(n)\Gamma$ on $X = G/\Gamma$ a polynomial sequence on X.

2.3.4 Quantitative equidistribution

Assume that $p(t)$ is a polynomial. Then, $p(n)$ can be expressed uniquely in the form

$$p(n) = \sum_{i=0}^{d} a_i n^i$$

for some real numbers a_i and $d \in \mathbb{N}$. For $N \in \mathbb{N}$, we define the smoothness norm[2]

$$\|e(p(n))\|_{C^\infty[N]} = \max_{1\leq i\leq d} (N^i \|a_i\|_{\mathbb{R}/\mathbb{Z}}). \tag{2.9}$$

A filtration on a Lie group G gives rise to a basis on its Lie algebra $\mathfrak{g}$, which is called a Mal'cev basis [43]. Mal'cev bases play an essential role in the theory of quantitative equidistribution on nilmanifolds. Firstly, we give the following definition:

Definition 2.3.5. *Let $X = G/\Gamma$ be a k-step nilmanifold with a rational filtration $G_\bullet = (G^{(i)})_{i\geq 0}$. Define $m = dim(G)$ and $m_i = dim(G^{(i)})$. A basis $(\xi_1, ..., \xi_m)$ of the associated Lie algebra $\mathfrak{g}$ over $\mathbb{R}$ is called a Mal'cev basis adapted to $G_\bullet$, if the following conditions are met:*
i) For each $0 \leq j \leq m-1$, $\mathfrak{h}_j = span(\xi_{h+1}, ..., \xi_m)$ is a Lie algebra ideal on $\mathfrak{g}$ and thus $H_j = \exp(\mathfrak{h}_j)$ is a normal Lie subgroup of G.
ii) For every $0 \leq i \leq k$, we have $G^{(i)} = H_{m-m_i}$.
iii) Each $b \in G$ can be uniquely written in the form $\exp(t_1\xi_1)\ldots\exp(t_m\xi_m)$ for $t_i \in \mathbb{R}$.
iv) The subgroup Γ consists precisely of those elements which, when written in the above form, have all $t_i \in \mathbb{Z}$.

[2] The definition of the smoothness norms is a bit different in [27]. There, the authors write the polynomials in the form $p(n) = \sum_{i=0}^{d} a_i \binom{n}{i}$ and define the smoothness norm using the same definition as (2.9) (the coefficients a_i are different). However, these definitions give two equivalent norms and, thus, all theorems can be stated for both norms, up to changes in the absolute constants.

Suppose that the element b is written in the form $\exp(t_1\xi_1)\cdots\exp(t_m\xi_m)$. The map $\psi : G \to \mathbb{R}^m$ defined by $\psi(b) = (t_1, ..., t_m)$ is a diffeomorphism from G to $\mathbb{R}^m$. The numbers $(t_1, ..., t_m)$ are called the coordinates of g with respect to the associated Mal'cev basis. If we consider the Euclidean metric on $\mathbb{R}^m$, we can construct a Riemannian metric d_G on G, whose value at the origin is equal to the Euclidean metric of $\mathbb{R}^m$ at the origin (of $\mathbb{R}^m$) composed with the inverse map ψ^{-1}. This metric is invariant under right translations and induces a metric d_X on $X = G/\Gamma$ defined by the relation:

$$d_X(g\Gamma, h\Gamma) = \inf\{d_G(b, b'), bg^{-1} \in \Gamma, b'h^{-1} \in \Gamma\}.$$

The metric used in [27] is slightly different than the one we consider here, but as the authors remark, these metrics are equivalent and all theorems hold as well by changing the absolute constants.

The sequence $(g(n)\Gamma)_{1\le n\le N}$ is said to be δ-equidistributed on the nilmanifold $X = G/\Gamma$ if and only if for any Lipschitz function $F : X \to \mathbb{C}$, we have that

$$\Big|\mathop{\mathbb{E}}_{1\le n\le N} F(g(n)\Gamma) - \int_X F d\mu_X\Big| \le \delta \, \|F\|_{\mathrm{Lip}(X)}$$

where

$$\|F\|_{\mathrm{Lip}(X)} = \|F\|_\infty + \sup_{x,y\in X,\ x\ne y} \frac{|F(x) - F(y)|}{d_X(x, y)}.$$

We now fix a k-step nilmanifold $X = G/\Gamma$, as well as a positive integer d. We equip it with the rational filtration $G_\bullet$ of degree dk that we defined above (after Definition 2.3.4), as well as a Mal'cev basis adapted to this filtration and the corresponding coordinate map $\psi : G \to \mathbb{R}^m$ (m is the dimension of G). Observe that under this filtration, we have that $G^{(d+1)} = G_2$ and property ii) in Definition 2.3.5 implies that $G_2 = H_{m-m_{d+1}}$. Thus, the Mal'cev basis induces an isometric identification of the horizontal torus $Z = G/G_2\Gamma$ with the torus $\mathbb{T}^{m-m_{d+1}}$ equipped with the standard metric.

Let $\pi : X \to Z$ denote the projection map and let χ be a horizontal character on G. Consider an element $b \in G$ with coordinates $(t_1, ..., t_m)$. Then, by properties iii) and iv) in Definition 2.3.5, we have that there is some $\vec{\ell} = (\ell_1, ..., \ell_{m-m_{d+1}}) \in \mathbb{Z}^{m-m_{d+1}}$ such that

$$\chi \circ \pi(b) = \ell_1 t_1 + \cdots + \ell_{m-m_{d+1}} t_{m-m_{d+1}}.$$

Thus, we get a character on the torus $\mathbb{T}^{m-m_{d+1}}$(written here with additive notation). We can then define the modulus $\|\chi\|$ of the character χ to be equal to

$$\left\|\vec{\ell}\right\| = |\ell_1| + \cdots + |\ell_{m-m_{d+1}}|. \tag{2.10}$$

If $v(n)$ is the polynomial sequence in (2.8) (recall that it is a polynomial sequence with respect to the filtration $G_\bullet$), then the sequence $\chi \circ \pi(v(n)\Gamma)$ is a polynomial sequence on the horizontal torus $Z \cong \mathbb{T}^{m-m_{d+1}}$. Indeed, if we denote $\psi(b_i) = (t_{i,1}, ..., t_{i,m})$, then a simple calculation shows that

$$\chi(\pi(v(n)\Gamma)) = \chi\big(\pi(b_1^{p_1(n)} \cdot ... \cdot b_k^{p_k(n)})\big) = \\ p_1(n)(\ell_1 t_{1,1} + \cdots + \ell_{m-m_{d+1}} t_{1,m-m_{d+1}}) + \cdots + p_k(n)(\ell_1 t_{k,1} + \cdots + \ell_{m-m_{d+1}} t_{k,m-m_{d+1}}),$$

which makes the fact that $\chi(\pi(v(n)\Gamma))$ is a polynomial sequence more evident.

The primary tool that we shall use is the following theorem of Green-Tao which describes the orbits of polynomial sequences in finite intervals. We present it in the case of our filtration $G_\bullet$, although the statement holds for any rational filtration. Some quantitative information (specifically relating to the concepts of quantitative rationality of Mal'cev bases) has been suppressed, since in our applications the nilmanifold will be fixed and the above condition on the Mal'cev bases is guaranteed if we take δ small enough.

Theorem J. *[27, Theorem 2.9] Let d be a non-negative integer, $X = G/\Gamma$ be a nilmanifold with G connected and simply connected and we equip the nilmanifold X with the Mal'cev basis adapted to the dk filtration $G_\bullet$ as above. Assume δ is a sufficiently small (depending only on X, d) parameter.*

Then, there exist a positive constant $C = C(X,d)$ with the following property: For every $N \in \mathbb{N}$, if $(v(n))_{n\in\mathbb{N}}$ is a polynomial sequence with respect to $G_\bullet$ such that the finite sequence $(v(n)\Gamma)_{1\le n\le N}$ is not δ-equidistributed, then for some non-trivial horizontal character χ (that depends on N and the sequence $v(n)$) of modulus $\|\chi\| \le \delta^{-C}$ we have

$$\|\chi(\pi(v(n)\Gamma))\|_{C^\infty(N)} \le \delta^{-C},$$

where π denotes the projection map from X to its horizontal torus.

In order to get a sense of how this theorem works, we refer the reader to Lemma 2.3.7 below which handles the case of polynomial sequences on tori. An important observation is that the constants do not depend on the length of the averaging interval and also depend only on the degree of the polynomial. Thus, we will be able to apply this theorem for polynomial sequences varying with N, as long as their degrees are kept constant.

2.3.5 Quantitative equidistribution in the abelian case

In the case of polynomial sequences on tori, their equidistribution properties are well understood. If the polynomial has rational non-constant coefficients, it is straightforward to check that the sequence of its fractional parts is periodic. For polynomials with at least one non-constant irrational coefficient, Weyl's theorem implies that the sequence is well-distributed modulo 1. In the case of Hardy field functions, we have a complete characterization of equidistribution modulo 1 due to Boshernitzan (Theorem E).

For the proofs of the results concerning primes, we will need quantitative information for the equidistribution of our sequences. This is provided through the means of discrepancy.

Definition 2.3.6. *Let $(u_n)_{1\le n\le N}$ be a finite sequence of real numbers $(u_n)_{1\le n\le N}$ and let $[a,b] \subseteq [0,1]$ be an interval. We define the* discrepancy *of the sequence u_n with respect to $[a,b]$ by*

$$\Delta_{[a,b]}(u_1,\dots,u_N) = \left| \frac{\big|\{n \in \{1,\dots,N\}\colon \{u_n\} \in [a,b]\}\big|}{N} - (b-a)\right|. \tag{2.11}$$

The discrepancy of a sequence is a quantitative measure of how close a sequence of real numbers is to being equidistributed modulo 1. For example, it is immediate that for an equidistributed sequence u_n, we have that

$$\lim_{N\to+\infty} \Delta_{[a,b]}(u_1,\dots,u_N) = 0,$$

for all $0 \le a \le b \le 1$. For an in-depth discussion on the concept of discrepancy and the more general theory of equidistribution on $\mathbb{T}$, we refer the reader to [37]. Our only tool will be an upper bound of Erdős and Turán on the discrepancy of a finite sequence. For a proof of this result, see [37, Chapter 2, Theorem 2.5].[3]

Theorem I (Erdős-Turán)**.** *There exists an absolute constant C, such that for any positive integer M and any Borel probability measure ν on $\mathbb{T}$, we have*

$$\sup_{A\subseteq\mathbb{T}} |\nu(A) - \lambda(A)| \le C\Big(\frac{1}{M} + \sum_{m=1}^{M} \frac{|\widehat{\nu}(m)|}{m}\Big),$$

where λ is the Lebesgue measure on $\mathbb{T}$ and the supremum is taken over all arcs A of $\mathbb{T}$.

In particular, specializing to the case that $\nu = N^{-1}\sum_{i=1}^{N} \delta_{\{u_i\}}$, where $u_1,\dots,u_N$ is a finite sequence of real numbers, we have

$$\Delta_{[a,b]}(u_1,\dots,u_N) \le C\Big(\frac{1}{M} + \sum_{m=1}^{M} \frac{1}{m}\Big|\frac{1}{N}\sum_{n=1}^{N} e(mu_n)\Big|\Big) \tag{2.12}$$

for all positive integers M and all $0 \le a \le b < 1$.

[3] In this book, the theorem is proven for measures of the form $\nu = \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}$, although the more general statement follows by noting that every Borel probability measure is a weak limit of measures of the previous form.

It is clear that in order to get the desired bounds on the discrepancy in our setting, we will need some estimates for exponential sums of Hardy field sequences in short intervals. Due to the Taylor approximation, this is morally equivalent to establishing estimates for exponential sums of polynomial sequences. There are several well-known estimates in this direction, the most fundamental of these being a result of Weyl that shows that an exponential sum along a polynomial sequence is small unless all non-constant coefficients of the polynomial satisfy a "major-arc" condition. In the case of strongly non-polynomial Hardy field functions, we will only need to study the leading coefficient of the polynomial in its Taylor approximation (in the proofs of the results involving primes), which will not satisfy such a major-arc condition.

Fortunately, Theorem J already provides a sufficiently strong estimate for tori. More precisely, we have the following:

Lemma 2.3.7. *Let $0<\delta<1$ and $d\in\mathbb{N}$. There exists a positive constant C depending only on d, such that if $p(x)=a_dx^d+\cdots+a_1x+a_0$ is a real polynomial that satisfies*

$$\Big|\frac{1}{N}\sum_{n=1}^{N}e(p(n))\Big|>\delta,$$

then, for every $1\leq k\leq d$, there exists $q\in\mathbb{Z}$ with $|q|\leq\delta^{-C}$, such that $N^k\left\|qa_k\right\|_{\mathbb{T}}\leq\delta^{-C}$.

Again, notice that there is no dependency of the constant on the length of the averaging interval, or on the implicit polynomial p apart from its degree).

2.3.6 Nilsequences and correlation sequences

An *s-step nilsequence* is a sequence of the form $F(g^nx)$, where $X=G/\Gamma$ is a s-step nilmanifold, $g\in G$, $x\in X$ and F is a continuous function on X. More precisely, we have the following definition for nilsequences in several variables.

Definition 2.3.8. *Let k,s be positive integers and let $X=G/\Gamma$ be a s-step nilmanifold. Assume that $g_1,\ldots,g_k$ are pairwise commuting elements of the group G, $F:X\to\mathbb{C}$ is a continuous function on X and $x\in X$. Then, the sequence*

$$\psi(n_1,\ldots,n_k)=F(g_1^{n_1}\cdot\ldots\cdot g_k^{n_k}x),\ \textit{where } n_1,\ldots,n_k\in\mathbb{Z},$$

*is called an s-*step nilsequence in *k-*variables*.*

Nilsequences arise naturally when studying the Gowers norms or the Host-Kra seminorms. A deep theorem of Green, Tao and Ziegler [28] implies that a sequence has a large Gowers norm (of degree s) if and only if it "correlates" with a s-step nilsequence.

The main tool that we will need is an approximation of general nilsequences by multi-correlation sequences in the ℓ^∞-sense. The following lemma is established in [18, Proposition 4.2].

Lemma 2.3.9. *Let k,s be positive integers and $\psi:\mathbb{Z}^k\to\mathbb{C}$ be a $(s-1)$-step nilsequence in k variables. Then, for every $\varepsilon>0$, there exists a system $(X,\mathcal{X},\mu,T_1,\ldots,T_k)$ and functions $F_1,\ldots,F_s$ on $L^\infty(\mu)$, such that the sequence $b(n_1,\ldots,n_k)$ defined by*

$$b(n_1,\ldots,n_k)=\int\prod_{j=1}^{s}(T_1^{\ell_jn_1}\cdot\ldots\cdot T_k^{\ell_jn_k})F_j\,d\mu,\ (n_1,\ldots,n_k)\in\mathbb{Z}^k$$

with $\ell_j=s!/j$ satisfies

$$\|\psi-b\|_{\ell^\infty(\mathbb{Z}^k)}\leq\varepsilon.$$

Remark. The definition of nilsequences used in [18] imposed that $x=\Gamma$ and that $\mathbf{n}\in\mathbb{N}^k$. However, their arguments generalize in a straightforward manner to the slightly more general setting that we presented above.

2.4 Background in number theory

In order to prove our results concerning averages over primes, we need some number theoretic input on the von Mangoldt function Λ. First of all, we need to show that studying averages over primes and averages weighted with the von Mangoldt function is the same. The following lemma is a standard consequence of the prime number theorem and the sparseness of prime powers (we use this argument in the proof of Corollary 2.4.3 below). For a proof, see, for instance, [30, Chapter 25].

Lemma 2.4.1. *For any bounded sequence $(a(n))_{n\in\mathbb{N}}$ in a normed space, we have*

$$\lim_{N\to+\infty}\Big\|\frac{1}{\pi(N)}\sum_{p\in\mathbb{P}:\, p\leq N} a(p)-\frac{1}{N}\sum_{n=1}^{N}\Lambda(n)a(n)\Big\|=0. \tag{2.13}$$

Therefore, in order to study ergodic averages along primes, we can replace them with the ergodic averages over $\mathbb{N}$ weighted by the function $\Lambda(n)$.

In our theorems, we will use a procedure to bound our ergodic averages over primes with the Gower norm of the weight $\Lambda_{W,b}$. For the modified von Mangoldt function, we will use the following deep theorem, which was recently established in [44].

Theorem K. *[44, Theorem 1.5] Let $\varepsilon>0$ and assume $L(N)$ is a positive sequence that satisfies the bounds $N^{\frac{5}{8}+\varepsilon}\leq L(N)\leq N^{1-\varepsilon}$. Let s be a fixed integer and let w be a positive integer. Then, if N is large enough depending on w, we have that*

$$\|\Lambda_{w,b}-1\|_{U^s(N,N+L(N)]}=o_w(1) \tag{2.14}$$

for every $1\leq b\leq W$ with $(b,W)=1$.

The celebrated theorem of Green-Tao on the uniformity of the von Mangoldt function established the previous theorem for the Gowers norm over the long interval $[1,N]$.

We will need to use the orthogonality of $\Lambda_{w,b}$ to polynomial phases in short intervals. This is an immediate consequence of the U^d uniformity in Theorem K in conjunction with an application of the van der Corput inequality (see Lemma 3.4.3 below) d times until the polynomial phase is eliminated. Alternatively, one can use Proposition 5.2.1 for a rotation on the torus $\mathbb{T}$ to carry out the reduction to Theorem K.[4] We omit its proof.

Lemma 2.4.2. *Let $L(N)$ be a positive sequence satisfying $N^{\frac{5}{8}+\varepsilon}\prec L(N)\prec N^{1-\varepsilon}$ for some $\varepsilon>0$. Then, we have that*

$$\max_{\substack{1\leq b\leq W\\ (b,W)=1}}\ \sup_{\substack{p\in\mathbb{R}[t]\\ \deg p=d}}\Big|\underset{N\leq n\leq N+L(N)}{\mathbb{E}}\big(\Lambda_{w,b}(n)-1\big)e(p(n))\Big|=o_w(1) \tag{2.15}$$

for every N large enough depending on w.

Remark. (*i*) The error term $o_w(1)$ depends on the degree d, but since this will be fixed in applications, we suppressed that dependence above.
(*ii*) Quantitative bounds for similar expressions (involving the more general class of nilsequences, as well) were the main focus in [44], though in that setting the authors used a different weight of the form $\Lambda-\Lambda^{\#}$, where $\Lambda^{\#}$ is a carefully chosen approximant for the von Mangoldt function arising from considerations of the (modified) Cramer random model for the primes.

Finally, we will also use a corollary of the Brun-Titchmarsh inequality to bound the contribution of bad residue classes in our ergodic averages by a constant term. For $q\geq 2$ and $(a,q)=1$, we denote by $\pi(x,q,a)$ the number of primes $\leq x$ that are congruent to a modulo q. Alternatively, one could also use the asymptotics for averages of Λ in short intervals that were established by Huxley [31], since $L(N)$ will be chosen to grow sufficiently fast in our applications.

[4] Evidently, both statements rely on similar complexity reduction arguments, though Proposition 5.2.1 is stated in much larger generality involving numerous polynomials.

Theorem B (Brun-Titchmarsh inequality)**.** *We have*

$$\pi(x+y,q,a)-\pi(x,q,a)\leq\frac{2y}{\phi(q)\log(\frac{y}{q})} \tag{2.16}$$

for every $x\geq y>q$.

While we referred to this as the Brun-Titchmarsh inequality, the previous theorem was established in [45] by Montgomery and Vaughan (prior results contained the term $2+o(1)$ in the numerator). We will need a variant of this theorem adapted to the von Mangoldt function, which follows easily from the previous theorem and a standard partial summation argument.

Corollary 2.4.3. *For every* $q\leq y\leq x$, *we have*

$$\sum_{\substack{x\leq n\leq x+y\\ n\equiv a\ (q)}}\Lambda(n)\leq\frac{2y\log x}{\phi(q)\log(\frac{y}{q})}+O\Big(\frac{y}{\log x}\Big)+O\big(x^{\frac{1}{2}}\log x\big).$$

Proof. Consider the function

$$\pi(x,q,a)=\sum_{\substack{1\leq n\leq x\\ n\equiv a\ (Q)}}1_{\mathbb{P}}(n)$$

as in the statement of Theorem B, defined for all $x\geq 3/2$. Let

$$\theta(x,q,a)=\sum_{\substack{1\leq n\leq x\\ n\equiv a\ (Q)}}1_{\mathbb{P}}(n)\log n,\quad \psi(x,q,a)=\sum_{\substack{1\leq n\leq x\\ n\equiv a\ (Q)}}\Lambda(n).$$

It is evident that

$$\Big|\theta(x,q,a)-\psi(x,q,a)\Big|\leq\sum_{p^k\leq x:\ p\in\mathbb{P},k\geq 2}\log p\leq x^{1/2}\log x, \tag{2.17}$$

since there are at most $x^{1/2}$ prime powers $\leq x$ and each one of them contributes at most $\log x$ in this sum. Now, we use summation by parts to deduce that

$$\theta(x+y,q,a)-\theta(x,q,a)=\sum_{\substack{x<n\leq x+y\\ n\equiv a\ (Q)}}1_{\mathbb{P}}(n)\log n+O(1)=\pi(x+y,q,a)\log(x+y)-$$
$$\pi(x,q,a)\log(x+1)+\sum_{\substack{x<n\leq x+y\\ n\equiv a\ (Q)}}\pi(n,q,a)\Big(\log n-\log(n+1)\Big)+O(1).$$

Using the inequalities $\log n-\log(n+1)\leq-(n+1)^{-1}$ and $\log(x+y)\leq\log x+y/x$, we deduce that

$$\theta(x+y,q,a)-\theta(x,q,a)\leq\log x\Big(\pi(x+y,q,a)-\pi(x,q,a)\Big)+\frac{\pi(x+y,q,a)y}{x}-$$
$$\sum_{\substack{x<n\leq x+y\\ n\equiv a\ (Q)}}\frac{\pi(n,q,a)}{n+1}+O(1).$$

Using the estimate $\pi(x,q,a)\ll\frac{x}{\phi(q)\log x}$ and Theorem B, we bound the sum in the previous expression by

$$\log x\frac{2y}{\phi(q)\log(\frac{y}{q})}+O\Big(\frac{(x+y)y}{\phi(q)x\log(x+y)}\Big)+O\Big(\sum_{\substack{x<n\leq x+y\\ n\equiv a\ (Q)}}\frac{1}{\phi(q)\log n}\Big)+O(1).$$

Since

$$\sum_{\substack{x<n\leq x+y \\ n\equiv a\ (Q)}} \frac{1}{\log n} \leq \int_{x}^{x+y} \frac{dt}{\log t} + O(1) = \frac{x+y}{\log(x+y)} - \frac{x}{\log x} + \int_{x}^{x+y} \frac{dt}{\log^2 t} + O(1) \leq$$

$$\frac{y}{\log x} + O(\frac{y}{\log^2 x}) + O(1),$$

we conclude that

$$\theta(x+y,q,a) - \theta(x,q,a) \leq \frac{2y\log x}{\phi(q)\log(\frac{y}{q})} + O(\frac{y}{\log x}) + O(1). \tag{2.18}$$

Consequently, if we combine (2.17) and (2.18), we arrive at

$$\psi(x+y,q,a) - \psi(x,q,a) \leq \frac{2y\log x}{\phi(q)\log(\frac{y}{q})} + O(\frac{y}{\log x}) + O(x^{\frac{1}{2}}\log x),$$

as was to be shown. □

Remark. We will apply this corollary for $q = W$ and $y \gg x^{5/8+\varepsilon}$. Note that for y in this range, the second error term can be absorbed into the first one.

Chapter 3

Joint ergodicity of Hardy field sequences

3.1 The main proposition

In this section, we state the principal result that we will prove in this chapter and which asserts that the Host-Kra factors of a given system are characteristic for the convergence of the averages (1.4). We will need to make some small reductions to the original problem and prove several lemmas. The proof is quite lengthy and heavy notationally. Thus, we also provide some examples that present the main ideas, while avoiding most of the technicalities.

Proposition 3.1.1. *Let $\mathcal{H}$ be a Hardy field that is closed under compostion and compostional inversion of functions, when defined. Assume that the functions $a_1, a_2, ..., a_k \in \mathcal{H}$ have polynomial growth and suppose that the following two conditions hold:*

i) The functions $a_1, ..., a_k$ dominate the logarithmic function $\log t$.

ii) The pairwise differences $a_i - a_j$ dominate the logarithmic function $\log t$ for any $i \neq j$.

Then, there exists a positive integer s depending only on the functions $a_1, ..., a_k$, such that for any measure preserving system (X, μ, T), functions $f_1 \in L^\infty(\mu)$ and $f_{2,N} ..., f_{k,N} \in L^\infty(\mu)$, all bounded by 1, with $f_1 \perp Z_s(X)$, the expression

$$\sup_{|c_n|\leq 1} \left\| \mathop{\mathbb{E}}_{1\leq n\leq N} c_n \, T^{\lfloor a_1(n)\rfloor} f_1 \cdot T^{\lfloor a_2(n)\rfloor} f_{2,N} \cdot ... \cdot T^{\lfloor a_k(n)\rfloor} f_{k,N} \right\|_{L^2(\mu)} \tag{3.1}$$

converges to 0, as $N \to +\infty$.

Remarks. i) It is possible to establish Proposition 3.1.1 under the weaker assumption that only the functions $a_1, a_1 - a_2, ..., a_1 - a_k$ dominate the logarithmic function, but this requires a few more details in the proof and is not required for the proof of Theorem 1.1.2.
ii) It may be possible to establish that the number s does not, in fact, depend on the functions $a_1, ..., a_k$, but it can be bounded by a function involving the number k of functions and the highest degree[1] d of the involved functions. However, we do not concern ourselves here with the optimal value of s. In particular, we will use polynomial expansions of the functions $a_1, ..., a_k$ with degrees very large compared to the number d, which means that any possible dependence on d will be lost in the proof.

It is obvious that Proposition 3.1.1 implies Theorem 1.1.2 (this follows from a standard telescoping argument). The reason that we work with sequences of functions and the bounded sequence c_n is because that will be helpful in some spots to absorb some of the error terms that will appear in the iterates and also allows us to "transform" the sequences in the iterates, so that we can reduce our problem to the case that the first sequence a_1 has some specific properties depending on the situation at hand. As an example, we claim that we only need to consider the case when the function $a_1(t)$ has maximal growth in the family $\{a_1, ..., a_k\}$. Indeed, suppose that this is not the case. Then, there exists a function a_i for some $i \in \{1, ..., k\}$ with $a_1 \prec a_i$. Without loss of generality, assume that the

[1] This means the smallest integer d, for which $a_i(t) \ll t^d$ for all $1 \leq i \leq k$.

function a_k has maximal growth rate. It is sufficient to show that for any sequence of functions g_N with $||g_N||_{L^\infty(\mu)} \leq 1$, we have

$$\lim_{N\to\infty} \underset{1\leq n\leq N}{\mathbb{E}} c_{n,N} \int g_N \; T^{\lfloor a_1(n)\rfloor} f_1 \cdot ... \cdot T^{\lfloor a_k(n)\rfloor} f_{k,N} \; d\mu = 0.$$

Then, we can choose the function g_N to be the conjugate of the average

$$\underset{1\leq n\leq N}{\mathbb{E}} c_{n,N} \int T^{\lfloor a_1(n)\rfloor} f_1 \cdot ... \cdot T^{\lfloor a_k(n)\rfloor} f_{k,N} \; d\mu$$

to get our claim. Composing with $T^{-\lfloor a_k(n)\rfloor}$ and applying the Cauchy-Schwarz inequality, it is sufficient to show that

$$\lim_{N\to+\infty} \sup_{|c_n|\leq 1} \Big\|\underset{1\leq n\leq N}{\mathbb{E}} c_n \; T^{-\lfloor a_k(n)\rfloor} g_N \cdot T^{\lfloor a_1(n)\rfloor - \lfloor a_k(n)\rfloor} f_1 \cdot ... \cdot T^{\lfloor a_{k-1}(n)\rfloor - \lfloor a_k(n)\rfloor} f_{k-1,N}\Big\|_{L^2(\mu)} = 0.$$

We can write $\lfloor a_i(n)\rfloor - \lfloor a_k(n)\rfloor = \lfloor a_i(n) - a_k(n)\rfloor + e_{i,n}$, where the errors $e_{i,n}$ take values in $\{0, \pm 1\}$. Using Lemma 3.2.1 below, the errors can be absorbed by the supremum outside the average and, therefore, the function that corresponds to f_1 is equal to $a_1 - a_k$, which now has maximal growth rate among the new family of functions. It is also easy to check that the new family satisfies the conditions of Proposition 3.1.1.

3.2 Some averaging lemmas

This notion of absorbing the errors that we described above can be made more precise by the next lemma.

Lemma 3.2.1. *Assume that the integers $e_{i,n,N}$ take values in a finite set S. Then, for any sequences $a_{i,N}$ of integers, complex numbers $c'_{n,N}$ bounded in magnitude by 1 and any 1-bounded functions $f_{i,N}$, we have*

$$\Big\|\underset{1\leq n\leq N}{\mathbb{E}} \; c'_{n,N} \; T^{a_{1,N}(n)+e_{1,n,N}} f_{1,N} \cdot ... \cdot T^{a_{k,N}(n)+e_{k,n,N}} f_{k,N}\Big\|_{L^2(\mu)} \ll_{k,S}$$
$$\sup_{|c_{n,N}|\leq 1} \; \sup_{||f_2||_\infty\leq 1,...,||f_k||_\infty\leq 1} \Big\|\underset{1\leq n\leq N}{\mathbb{E}} \; c_{n,N} \; T^{a_{1,N}(n)} f_{1,N} \cdot T^{a_{2,N}(n)} f_2 \cdot ... \cdot T^{a_{k,N}(n)} f_k\Big\|_{L^2(\mu)}.$$

As a consequence, there exist 1-bounded functions $f'_{i,N}$, such that the original expression is bounded by a constant multiple of the quantity

$$\sup_{|c_{n,N}|\leq 1} \Big\|\underset{1\leq n\leq N}{\mathbb{E}} \; c_{n,N} \; T^{a_{1,N}(n)} f_{1,N} \; T^{a_{2,N}(n)} f'_{2,N} \cdot ... \cdot T^{a_{k,N}(n)} f'_{k,N}\Big\|_{L^2(\mu)} + o_N(1).$$

Proof. We partition the integers n into a finite number of sets, in which all the quantities $e_{i,n,N}$ are constant (as n varies). There are at most $|S|^k$ such sets. If $A_1, ..., A_{|S|^k}$ are these sets, then we have

$$\Big\|\underset{1\leq n\leq N}{\mathbb{E}} c'_{n,N} \; T^{a_{1,N}(n)+e_{1,n,N}} f_{1,N} \cdot ... \cdot T^{a_{k,N}(n)+e_{k,n,N}} f_{k,N}\Big\|_{L^2(\mu)} \leq$$
$$\sum_{i=1}^{|S|^k} \Big\|\frac{1}{N} \sum_{n\in A_i} \; c'_{n,N} \; T^{a_{1,N}(n)+e_{1,n,N}} f_{1,N} \cdot ... \cdot T^{a_{k,N}(n)+e_{k,n,N}} f_{k,N}\Big\|_{L^2(\mu)} \leq$$
$$|S|^k \max_{1\leq i\leq |S|^k} \Big\|\frac{1}{N} \sum_{1\leq n\leq N} \; c'_{n,N} \mathbf{1}_{A_i}(n) \; T^{a_{1,N}(n)} f_{1,N} \cdot ... \cdot T^{a_{k,N}(n)+e_{k,n,N}-e_{1,n,N}} f_{k,N}\Big\|_{L^2(\mu)} \leq$$
$$|S|^k \sup_{|c_{n,N}|\leq 1} \; \sup_{||f_2||_\infty\leq 1,...,||f_k||_\infty\leq 1} \Big\|\underset{1\leq n\leq N}{\mathbb{E}} c_{n,N} \; T^{a_{1,N}(n)} f_{1,N} \; T^{a_{2,N}(n)} f_2 \cdot ... \cdot T^{a_{k,N}(n)} f_k\Big\|_{L^2(\mu)},$$

which is the required result. In the second to last relation, we composed with $T^{-e_{1,n,N}}$, because $e_{1,n,N}$ is constant when n is restricted to the set A_i. □

Remark. In the following sections, we will encounter situations where we have some error terms in the iterates. The above lemma is not applied verbatim to all cases below. However, the reasoning presented above (i.e. partitioning into sets where the error sequences are constant) can be applied directly every time to remove these error terms. In particular, we can also show (using the same arguments) that a similar statement holds for double averages, that is, if I_r are a sequence of intervals with lengths going to infinity, d is a natural number and the error terms $e_{i,n,R}$ take values on a finite set S of integers, then

$$\mathop{\mathbb{E}}_{1\leq r\leq R}\Big\|\mathop{\mathbb{E}}_{n\in I_r} c'_{n,R}\, T^{a_{1,R}(n)+e_{1,n,R}}f_{1,R}\cdot\ldots\cdot T^{a_{k,R}(n)+e_{k,n,R}}f_{k,R}\Big\|^d_{L^2(\mu)} \ll_{S,k,d}$$

$$\sup_{\|f_2\|_\infty\leq 1,\ldots,\|f_k\|_\infty\leq 1}\ \mathop{\mathbb{E}}_{1\leq r\leq R}\ \sup_{|c_{n,R}|\leq 1}\ \Big\|\mathop{\mathbb{E}}_{n\in I_r} c_{n,R}\, T^{a_{1,R}(n)}f_{1,R}\cdot T^{a_{2,R}(n)}f_2\ldots\cdot T^{a_{k,R}(n)}f_k\Big\|^d_{L^2(\mu)},$$

where we also use the Hölder inequality (which gives dependence on the exponent d in the implicit constants). Therefore, instead of using the same argument repeatedly, we will cite this lemma in such instances and add a comment when a modified version is required.

The second lemma implies that we can bound our averages over the long intervals $[1, N]$, if we can obtain bounds over short inervals of the form $[r, r + L(r)]$ for some positive sub-linear function $L(t)$. Due to several obstructions in the proof, we will have to average over the parameter r as well, meaning that we will have to use a double averaging scheme in the proof.

Lemma 3.2.2. *Let d be a positive integer and consider a two-parameter sequence $\big(A_{R,n}\big)_{R,n\in\mathbb{N}}$ in a normed space such that $\|A_{R,n}\|\leq 1$ for all possible choices of $R, n\in\mathbb{N}$. Let $L(t)\in\mathcal{H}$ be an eventually positive function such that $1\prec L(t)\prec t$ and assume that*

$$\limsup_{R\to+\infty}\mathop{\mathbb{E}}_{1\leq r\leq R}\Big\|\mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} A_{R,n}\Big\|^d\leq C$$

for some $C > 0$. Then, we also have

$$\limsup_{R\to+\infty}\Big\|\mathop{\mathbb{E}}_{1\leq n\leq R} A_{R,n}\Big\|\leq C^{1/d}.$$

Proof. Combining the power mean inequality and the triangle inequality, we can easily deduce that

$$\mathop{\mathbb{E}}_{1\leq r\leq R}\Big\|\mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} A_{R,n}\Big\|^d\geq\Big\|\mathop{\mathbb{E}}_{1\leq r\leq R}\big(\mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} A_{R,n}\big)\Big\|^d.$$

Therefore, our result will follow if we show that

$$\Big\|\mathop{\mathbb{E}}_{1\leq r\leq R}\big(\mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} A_{R,n}\big)-\mathop{\mathbb{E}}_{1\leq n\leq R} A_{R,n}\Big\| = o_R(1).$$

Let u be the compositional inverse of the function $t + L(t)$. Our assumptions on the Hardy field $\mathcal{H}$ imply that $u\in\mathcal{H}$. In addition, it is easy to check that $\lim\limits_{t\to+\infty} u(t)/t = 1$. Now, we have

$$\mathop{\mathbb{E}}_{1\leq r\leq R}\big(\mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} A_{R,n}\big)=\frac{1}{R}\Big(\sum_{n=1}^{R} p_R(n)A_{R,n}+\sum_{n=R+1}^{R+L(R)} p_R(n)A_{R,n}\Big)$$

for some real numbers $p_R(n)$. Assuming that n (and thus R) is sufficiently large (so that $u(n)$ is positive) we can calculate $p_R(n)$ to be equal to

$$p_R(n)=\frac{1}{L(\lfloor u(n)\rfloor)+1}+\cdots+\frac{1}{L(n)+1}+o_n(1),$$

since the number $A_{R,n}$ appears on the average $\mathop{\mathbb{E}}\limits_{r\leq n\leq r+L(r)}$ if and only if $u(n)\leq r\leq n$. Note that $p_R(n)$ is actually independent of R (for n large enough) and therefore, we will denote it simply as $p(n)$ from now on. We claim that

$$\lim_{n\to+\infty} p(n)=1. \tag{3.2}$$

Let us first see how this finishes the proof. Since for n large enough we must have $p(n) \leq 2$, we can easily deduce that

$$\frac{1}{R}\sum_{n=R+1}^{R+L(R)} p(n)A_{R,n} = o_R(1).$$

Here, we used the fact that $L(t) \prec t$. In addition, we have

$$\Big\|\frac{1}{R}\sum_{n=1}^{R} p(n)A_{R,n} - \frac{1}{R}\sum_{n=1}^{R} A_{R,n}\Big\| \leq \frac{1}{R}\sum_{n=1}^{R} |p(n)-1|,$$

which is also $o_R(1)$. Combining the above we reach the desired conclusion.

In order to establish (3.2), we observe that $L(t)$ is eventually strictly increasing, and therefore, we can easily get

$$\int\limits_{\lfloor u(n)\rfloor}^{n+1} \frac{1}{L(t)+1}\, dt \leq p(n) \leq \int\limits_{\lfloor u(n)\rfloor -1}^{n} \frac{1}{L(t)+1}\, dt.$$

Thus, it suffices to show that the integrals on both sides of the above inequality converge to 1. It is straightforward to check that each of these integrals is $o_n(1)$ close to the integral

$$I_n = \int\limits_{u(n)}^{n} \frac{1}{L(t)+1}\, dt.$$

Therefore, we only need to prove that $I_n \to 1$. Using the mean value theorem, we can find a real number $h_n \in [u(n), n]$ such that,

$$I_n = \frac{n-u(n)}{L(h_n)+1} = \frac{L(u(n))}{L(h_n)+1}.$$

The last equality follows easily from the definition of u. Since L is eventually strictly increasing, we conclude that I_n is smaller than $L(u(n))/(L(u(n))+1) \leq 1$. In addition, we also have

$$I_n \geq \frac{L(u(n))}{L(n)+1}.$$

The result follows if we show (note that the function u^{-1} is onto in a half line of $\mathbb{R}$)

$$\lim_{t\to+\infty} \frac{L(t)}{L(u^{-1}(t))+1} = 1.$$

However,

$$\frac{L(t)}{L(u^{-1}(t))+1} = \frac{L(t)}{L(t+L(t))+1} = \frac{L(t)}{L(t+L(t))} + o_t(1).$$

Using the mean value theorem, we can write

$$L(t+L(t)) = L(t) + L(t)L'(x_t)\ ,$$

where $x_t \in [t, t+L(t)]$. Thus,

$$\frac{L(t+L(t))}{L(t)} = 1 + L'(x_t) = 1 + o_t(1)\ ,$$

since $L'(t) \ll L(t)/t \prec 1$. The result follows. □

Finally, we will use a change of variables lemma very similar to [13, Lemma 5.1], which can be proven similarly by a standard partial summation argument. It will allows us to bound quantities evaluated along sub-linear sequences with the same quantity evaluated along the sequence $a(n) = n$, which will simplify some expressions significantly.

Lemma 3.2.3. *Let $(V_R(n))_{n,R\in\mathbb{N}}$ be a 1-bounded, two-parameter sequence of vectors in a normed space and let $a \in \mathcal{H}$ satisfy the growth condition $t^\delta \prec a(t) \prec t$. Then, we have*

$$\limsup_{R\to+\infty} \Big\| \mathop{\mathbb{E}}_{1\leq n\leq R} V_R(\lfloor a(n)\rfloor)\Big\| \ll_a \limsup_{R\to+\infty} \Big\| \mathop{\mathbb{E}}_{1\leq n\leq R} V_R(n)\Big\|.$$

3.3 Overview of the proof and examples

Our main objective is to reduce our problem to the study of ergodic averages of some variable polynomials. Therefore, we will first study asymptotic bounds for certain polynomial families in the next section, since they will be required for the proof of Proposition 3.1.1. This will rely on the van der Corput inequality and an induction argument on the complexity of the family. The PET induction argument has been carried out for polynomials in the literature, although our setting necessitates that the bounds are finitary in nature (and, thus, more complicated). Following that, we need to show that Hardy field sequences can be simultaneously approximated by Taylor polynomials, which we will also need to satisfy several nice properties. This will be the content of Section 3.5 In Section 3.6, we will establish bounds for Hardy sequences of a specific form, namely when the involved functions are a sum of a sub-linear function and a polynomial. In Section 3.7, we shall reduce the problem to the case of functions that are a sum of a sub-linear function and a polynomial and finish the proof.

The main idea is that we can approximate the given Hardy functions by Taylor polynomials (possibly constant) in suitable smaller intervals (with lengths going to infinity). We shall reduce our problem to proving a statement of the form

$$\lim_{R\to+\infty} \underset{1\leq r\leq R}{\mathbb{E}} \Big\| \underset{n\in I_r}{\mathbb{E}}\, c_{n,R}\, T^{\lfloor p_{1,r}(n)\rfloor} f_{1,r}\cdot \ldots \cdot T^{\lfloor p_{k,r}(n)\rfloor} f_{k,r}\Big\|_{L^2(\mu)}^{2^t} = 0, \tag{3.3}$$

where the iterates are variable polynomials and $f_{1,r}$ has the form

$$f_{1,r} = f_1 \cdot T^{\lfloor b_1(r)\rfloor} h_1 \cdot \ldots \cdot T^{\lfloor b_\ell(r)\rfloor} h_\ell$$

for sub-linear functions $b_1, ..., b_\ell$ and $h_1, ..., h_\ell \in L^\infty(\mu)$.

After this reduction, we bound the innermost average using the results from Section 3.4. More precisely, we claim that the inner average can be bounded by a quantity of the form

$$\underset{\mathbf{m}\in[-M,M]^t}{\mathbb{E}} \Big| \int T^{\lfloor q_1(r,\mathbf{m})\rfloor} g_{r,1}\cdot \ldots \cdot T^{\lfloor q_\ell(r,\mathbf{m})\rfloor} g_{r,\ell}\, d\mu\Big|$$

plus some small error terms, where M is a finite integer (independent from the rest of our parameters) and all the functions $g_{r,i}$ are either $\tilde{f}_r$ or $\overline{\tilde{f}_r}$. In addition, the functions $q_i(r,\mathbf{m})$ in the iterates are such that, for (almost all) $\mathbf{m}\in\mathbb{Z}^l$, they can be written as a sum of a sublinear function plus a polynomial, which is the special case that we discussed above. Thus, taking first the limit $R\to+\infty$ to use the bounds established in the special case and then taking the limits $M\to+\infty$, we shall reach our conclusion.

3.3.1 Some examples

a) Whenever we use $\ll$ without indices in this example, we imply that the constants are absolute. Assume that $a(t) = t\log t + \log^3 t$, $b(t) = t\log t$ and $c(t) = \sqrt{t}$. We want to show that there exists $s\in\mathbb{N}$, such that, if $|||f|||_s = 0$, then

$$\underset{1\leq n\leq N}{\mathbb{E}} T^{\lfloor n\log n + \log^3 n\rfloor} f \cdot T^{\lfloor n\log n\rfloor} g_1 \cdot T^{\lfloor\sqrt{n}\rfloor} g_2$$

converges to 0 in L^2 as $N\to+\infty$. Here, g_1 and g_2 are arbitrary 1-bounded functions in $L^\infty(\mu)$. In view of Lemma 3.2.2, it suffices to show that

$$\underset{1\leq r\leq R}{\mathbb{E}} \Big\| \underset{r\leq n\leq r+L(r)}{\mathbb{E}} T^{\lfloor n\log n+\log^3 n\rfloor} f\cdot T^{\lfloor n\log n\rfloor} g_1\cdot T^{\lfloor\sqrt{n}\rfloor} g_2\Big\|_{L^2(\mu)}^{2^d} = \underset{1\leq r\leq R}{\mathbb{E}} A_r \tag{3.4}$$

converges to 0 as $R\to+\infty$, for some sub-linear function $L(t)\in\mathcal{H}$ and an integer d, both of which we will choose later.

Step 1: Reduction to averages of variable polynomials.

We observe that

$$A_r = \Big\|\mathop{\mathbb{E}}_{0\le h\le L(r)} T^{\lfloor (r+h)\log(r+h)+\log^3(r+h)\rfloor} f\cdot T^{\lfloor (r+h)\log(r+h)\rfloor} g_1\cdot T^{\lfloor \sqrt{r+h}\rfloor} g_2\Big\|_{L^2(\mu)}^{2^d}.$$

Now, we can use the Taylor expansion to write

$$(r+h)\log(r+h) = -\frac{h^3}{6x_h^2}+\frac{h^2}{2r}+h(\log r+1)+r\log r, \quad \text{for some} \quad x_h\in[r,r+h],$$

and

$$\sqrt{r+h} = -\frac{h^2}{8(x_h')^{3/2}}+\frac{h}{2\sqrt{r}}+\sqrt{r}, \quad \text{for some} \quad x_h'\in[r,r+h],$$

for every $0\le h\le L(r)$. Since

$$\Big|\frac{h^3}{6x_h^2}\Big|\le\frac{L(r)^3}{r^2}$$

and

$$\Big|\frac{h^2}{8(x_h')^{3/2}}\Big|\le\frac{L^2(r)}{8r^{3/2}},$$

we conclude that these two last terms are both $o_r(1)$, provided that we choose the function $L(t)$ to satisfy $L(t)\prec t^{2/3}$. We also choose $L(t)\succ t^{1/2}$, so that both the 2-degree term in the expansion of $(r+h)\log(r+h)$ and the 1-degree term in the expansion of $\sqrt{r+h}$ are not bounded (for h taking values in the range $[0,L(r)]$). In addition, under the above assumptions, we can also show that

$$\max_{0\le h\le L(r)}|\log^3(r+h)-\log^3(r)| = o_r(1)$$

using the mean-value theorem. Therefore, we have[2]

$$A_r\simeq \Big\|\mathop{\mathbb{E}}_{0\le h\le L(r)} T^{\left\lfloor \frac{h^2}{2r}+h(\log r+1)+r\log r+\log^3 r\right\rfloor} f\cdot T^{\left\lfloor \frac{h^2}{2r}+h(\log r+1)+r\log r\right\rfloor} g_1\cdot T^{\left\lfloor \frac{h}{2\sqrt{r}}+\sqrt{r}\right\rfloor} g_2\Big\|_{L^2(\mu)}^{2^d} =$$
$$\Big\|\mathop{\mathbb{E}}_{0\le h\le L(r)} T^{\left\lfloor \frac{h^2}{2r}+h(\log r+1)+r\log r\right\rfloor}(g_1\cdot T^{\lfloor \log^3 r\rfloor} f)\cdot T^{\left\lfloor \frac{h}{2\sqrt{r}}+\sqrt{r}\right\rfloor} g_2\Big\|_{L^2(\mu)}^{2^d}, \quad (3.5)$$

which is an average where the iterates are polynomials in h. The fact that the $o_r(1)$ terms can be discarded follows from Lemma 3.2.1 and will be explained in depth in the formal proof. Note that the iterates have now become polynomials in the variable h.

Remark. In the proof of Proposition 3.1.1 in Section 3.7, we will choose the function $L(t)$ in order to have a common polynomial expansion as above. Although in this example this is easily done by hand, this will be accomplished in the general case using some lemmas and propositions that are proven.

We will use the van der Corput inequality (Lemma 3.4.4):

$$\Big|\mathop{\mathbb{E}}_{1\le n\le N} a_n\Big|^{2^d}\ll_d \frac{1}{M}+\mathop{\mathbb{E}}_{|m|\le M}\Big|\mathop{\mathbb{E}}_{1\le n\le N}\langle a_{n+m},a_m\rangle\Big|^{2^{d-1}}+o_N(1)\ ,$$

which holds as long as $M=o(N)$.

We will deal with a simpler case here, since (3.5) requires many applications of the van der Corput inequality and the estimates are quite complicated. We shall find a bound for the average

$$\mathop{\mathbb{E}}_{1\le r\le R}\Big\|\mathop{\mathbb{E}}_{0\le h\le L(r)} T^{\left\lfloor \frac{h^2}{2r}\right\rfloor} f_r\Big\|^4 = \mathop{\mathbb{E}}_{1\le r\le R} A_r^4\ ,$$

where $f_r = g_1\cdot T^{\lfloor \log^3(r)\rfloor} f$.

[2]In this example, we split and combine the integer parts freely, which is not true in general. In our main proof, we explain this argument using Lemma 3.2.1.

Step 2: A change of variables trick and bounds for the polynomial averages

First of all, we can write $h = k\lfloor\sqrt{2r}\rfloor + s$, where the integers k, s satisfy $0 \le k \le L(r)/\lfloor\sqrt{2r}\rfloor$ and $0 \le s \le \lfloor\sqrt{2r}\rfloor - 1$. Then, we have

$$\frac{h^2}{2r} = \frac{k^2\lfloor\sqrt{2r}\rfloor^2}{2r} + \frac{2k\lfloor\sqrt{r}\rfloor s}{2r} + \frac{s^2}{2r}.$$

Note that

$$\Big|\frac{k^2\lfloor\sqrt{2r}\rfloor^2}{2r} - k^2\Big| \le 2k^2\frac{\{\sqrt{2r}\}}{\sqrt{2r}} \le 2\frac{L^2(r)}{\lfloor\sqrt{2r}\rfloor^2\sqrt{2r}}.$$

If we choose $L(t)$ to satisfy the additional hypothesis $L(t) \prec t^{3/4}$ (which is vacuous since we already have the restriction $L(t) \prec t^{2/3}$), then we get that the above quantity is $o_r(1)$. In this example, we can take $L(t) = t^{3/5}$ as our sub-linear function (observe that all of the restrictions we imposed above are satisfied). Therefore, we can use the power mean inequality to deduce that

$$A_r^4 \le \underset{0\le s\le\lfloor\sqrt{2r}\rfloor-1}{\mathbb{E}} \Big\| \underset{1\le k\le \frac{L(r)}{\lfloor\sqrt{2r}\rfloor}}{\mathbb{E}} T^{\lfloor k^2+p_{s,r}(k)\rfloor} f_r \Big\|^4 \tag{3.6}$$

for some linear polynomials $p_{s,r}(k)$. Denote by $A_{s,r}$ the innermost average in the above relation.

We fix a positive integer parameter M. Applying the van der Corput inequality twice, we deduce that

$$A_{s,r}^4 \ll \frac{1}{M} + \underset{|m_1|,|m_2|\le M}{\mathbb{E}} \Big| \int \bar{f}_r \cdot T^{2m_1m_2} f_r \, d\mu \Big| + o_r(1),$$

where the implied constant is absolute (and, in particular, independent of M). We omitted the routine computations here (the general case is more complicated than this and is handled in Section 3.4). This bound holds regardless of the choice of the polynomial $p_{s,r}(k)$ and the only thing that influences this bound is the highest order coefficient (which is 1 for all values of s, r).

Using this bound in (3.6) we deduce that

$$A_r^4 \ll \frac{1}{M} + \underset{|m_1|,|m_2|\le M}{\mathbb{E}} \Big| \int \overline{(g_1 \cdot T^{\lfloor \log^3 r\rfloor} f)} \cdot T^{2m_1m_2}(g_1 \cdot T^{\lfloor \log^3 r\rfloor} f) \, d\mu \Big| + o_r(1).$$

Remark. It is now clear that we need the outer average over r in the beginning because this last bound is only useful on average. More generally, this problem will always appear if some of the functions on the iterates or the differences of two functions in the iterates are sub-fractional and some others are super-fractional. The sub-fractional functions are essentially constant in the short intervals and, thus, the bounds we get will depend on these constant values. We average again in order to handle the sub-fractional functions.

Therefore, the quantity in (3.4) is $\ll$

$$\frac{1}{M} + \underset{1\le r\le R}{\mathbb{E}} \underset{|m_1|,|m_2|\le M}{\mathbb{E}} \Big| \int \overline{(g_1 \cdot T^{\lfloor \log^3 r\rfloor} f)} \cdot T^{2m_1m_2}(g_1 \cdot T^{\lfloor \log^3 r\rfloor} f) \, d\mu \Big| + o_R(1) =$$
$$\frac{1}{M} + \underset{|m_1|,|m_2|\le M}{\mathbb{E}} \underset{1\le r\le R}{\mathbb{E}} \Big| \int (\bar{g}_1 \cdot T^{2m_1m_2} g_1) \cdot T^{\lfloor \log^3 r\rfloor}(\bar{f} \cdot T^{2m_1m_2} f) \, d\mu \Big| + o_R(1). \tag{3.7}$$

Remark. In the proof of the general case, instead of the sub-linear function $\lfloor \log^3(r)\rfloor$ in the iterates in (3.7), we may also have functions of the form $\lfloor u(r)\rfloor^k$, where $u \in \mathcal{H}$ is a sub-linear function and $k \in \mathbb{Z}^+$ (like $\lfloor\sqrt{r}\rfloor^3$ and $\lfloor r^{2/3}\rfloor^5$). For instance, assume we want to study the limit of the averages

$$\underset{1\le n\le N}{\mathbb{E}} T^{\lfloor \sqrt{n}+n^3\rfloor} f \cdot T^{\lfloor\sqrt{n}\rfloor} g.$$

Using Lemma 3.2.2, it suffices to show that

$$\underset{1\le r\le R}{\mathbb{E}} \Big\| \underset{0\le h\le L(r)}{\mathbb{E}} T^{\lfloor \sqrt{r+h}+(r+h)^3\rfloor} f \cdot T^{\lfloor\sqrt{r+h}\rfloor} g \Big\|_{L^2(\mu)}^{2^d}$$

for some $d \in \mathbb{N}$ and some sub-linear function $L(t) \in \mathcal{H}$. If we choose $L(t)$ appropriately, then we can write

$$\sqrt{r+h} = \sqrt{r} + \frac{h}{2\sqrt{r}} + o_r(1)$$

for $0 \le h \le L(r)$. Now, using the change of variables $h = k\lfloor 2\sqrt{r} \rfloor + s$, we observe that the leading coefficient of the polynomial $(r+h)^3$ in the iterates becomes $\lfloor 2\sqrt{r} \rfloor^3$. If we proceed similarly as in step 2 above using repeated applications of the van der Corput inequality, we will arrive at a similar bound as the one in (3.7), but now the term $\lfloor 2\sqrt{r} \rfloor^3$ will appear in the iterates.

In order to combat this situation, we need another intermediate step in our proof (this is Step 7 in Section 3.7). We shall use Lemma 3.2.3 that allows us to replace the sub-linear function $2\sqrt{r}$ by the identity function $a(r) = r$. As an example, suppose we want to bound the limit of the averages

$$\mathop{\mathbb{E}}_{1 \le r \le R} T^{\lfloor \sqrt{r} \rfloor} f \cdot T^{\lfloor \sqrt{r} \rfloor^3 + \lfloor r^{2/5} \rfloor} g$$

as $R \to +\infty$. We rewrite this expression as a function of $\sqrt{r}$

$$\mathop{\mathbb{E}}_{1 \le r \le R} T^{\lfloor \sqrt{r} \rfloor} f \cdot T^{\lfloor \sqrt{r} \rfloor^3 + \lfloor (\sqrt{r})^{4/5} \rfloor} g.$$

Then, we can prove that

$$\limsup_{R \to +\infty} \big\| \mathop{\mathbb{E}}_{1 \le r \le R} T^{\lfloor \sqrt{r} \rfloor} f \cdot T^{\lfloor \sqrt{r} \rfloor^3 + \lfloor (\sqrt{r})^{4/5} \rfloor} g \big\|_{L^2(\mu)} \le C \limsup_{R \to +\infty} \big\| \mathop{\mathbb{E}}_{1 \le r \le R} T^r f \cdot T^{r^3 + \lfloor r^{4/5} \rfloor} g \big\|_{L^2(\mu)}$$

for some positive real number C. Now the functions in the iterates are sub-linear functions and polynomials, which we are now able to handle (this is the content of Section 3.6).

Step 3: Dealing with the sub-linear function.

In this step, we show that the quantity in (3.7) goes to 0, if we take $R \to +\infty$ and then $M \to +\infty$. While steps 1 and 2 of this example correspond to parts of the proof in Sections 3.4.5 and 3.7, this step corresponds to the proofs in Section 3.6.

We observe that the function $\log^3(r)$ in the iterates is a sub-linear function. We will show that

$$\lim_{R \to +\infty} \mathop{\mathbb{E}}_{1 \le r \le R} \Big| \int (\bar{g} \cdot T^{2m_1 m_2} g) \cdot T^{\lfloor \log^3 r \rfloor} (\bar{f} \cdot T^{2m_1 m_2} f) d\mu \Big| \ll \|\bar{f} \cdot T^{2m_1 m_2} f\|_3. \tag{3.8}$$

In addition, the implicit constants do not depend on m_1, m_2. Assuming that (3.8) holds, we take the limit as $M \to +\infty$ (this can be done because all implied asymptotic constants do not depend on m_1, m_2) and we need to show that

$$\lim_{M \to +\infty} \mathop{\mathbb{E}}_{|m_1|, |m_2| \le M} \|\bar{f} \cdot T^{2m_1 m_2} f\|_3 = 0.$$

Applying the Hölder inequality, we are left with showing that

$$\lim_{M \to +\infty} \mathop{\mathbb{E}}_{|m_1|, |m_2| \le M} \|\bar{f} \cdot T^{2m_1 m_2} f\|_3^8 = 0.$$

Using the definition of the Host-Kra seminorms, this relation reduces to an ergodic average with polynomial iterates, which is well known to converge to 0 under our hypothesis on the function f (namely, that $\|f_1\|_s = 0$ for some suitable $s \in \mathbb{N}$).

We now establish (3.8). It suffices to show that

$$\lim_{R \to +\infty} \mathop{\mathbb{E}}_{1 \le r \le R} \Big| \int g \cdot T^{\lfloor \log^3 r \rfloor} f d\mu \Big| \ll \|f\|_{3,T}$$

for any 1-bounded functions f and g, where the implied constant is absolute. We square the above expression and apply the Cauchy-Schwarz inequality to bound it by

$$\mathop{\mathbb{E}}_{1 \le r \le R} \int G \cdot S^{\lfloor \log^3(r) \rfloor} F \, d(\mu \times \mu),$$

where $F := \overline{f} \otimes f$, $G := \overline{g} \otimes g$ and $S := T \times T$. Then, (3.8) follows if we show

$$\Big\| \mathop{\mathbb{E}}_{1 \le r \le R} S^{\lfloor \log^3(r) \rfloor} F \Big\|_{L^2(\mu \times \mu)} \ll \|\!| f |\!\|_{3,T}^2.$$

We use Lemma 3.2.2 once more: it suffices to show that

$$\limsup_{r \to +\infty} \Big\| \mathop{\mathbb{E}}_{r \le n \le r+L(r)} S^{\lfloor \log^3(n) \rfloor} F \Big\|_{L^2(\mu \times \mu)} \ll \|\!| f |\!\|_{3,T}^2 ,$$

where $L(t) \in \mathcal{H}$ is sub-linear. Using the Taylor expansion, we can write

$$\log^3(r+h) = \log^3(r) + \frac{3\log^2 r}{r} h - \frac{6\log x_h - 3\log^2 x_h}{2x_h^2} h^2,$$

where $0 \le h \le L(r)$ and $x_h \in [r, r+h]$. If we choose the function $L(t)$ so that

$$\frac{t}{\log^2 t} \prec L(t) \prec \frac{t}{\log t},$$

we can then deduce that the last term in the above expansion is $o_r(1)$. Our problem reduces to

$$\limsup_{r \to +\infty} \Big\| \mathop{\mathbb{E}}_{0 \le h \le L(r)} S^{\left\lfloor \log^3(r) + \frac{3\log^2 r}{r} h \right\rfloor} F \Big\|_{L^2(\mu \times \mu)} \ll \|\!| f |\!\|_{3,T}^2.$$

We have again reduced our problem to finding a bound for an ergodic average with (variable) polynomials. In order to finish the proof, we work similarly as in the previous steps, using the change of variables trick and one application of the van der Corput inequality (we also need to use the inequality $\|\!| F |\!\|_{2,T\times T} \le \|\!| f |\!\|_{3,T}^2$).

b) In this second example we describe the strategy that will be used in the special case that we discussed above, that is when our functions are sums of sub-linear functions and polynomials. This case is covered in full generality in Section 3.6. We consider the triplet of functions in $\mathcal{H}$ $(t + \log^3 t, t, \log^2 t)$ and we shall show that there exists $s \in \mathbb{N}$ so that, if $\|\!| f |\!\|_s = 0$, then

$$\mathop{\mathbb{E}}_{1 \le n \le N} T^{\lfloor n + \log^3 n \rfloor} f \cdot T^n g_1 \cdot T^{\lfloor \log^2 n \rfloor} g_2$$

converge to 0 in mean (g_1, g_2 are again arbitrary 1-bounded functions).

Step 1: Reducing to the case when all iterates have sub-linear growth.

We start by using Lemma 3.2.2 to reduce our problem to

$$\limsup_{R \to +\infty} \mathop{\mathbb{E}}_{1 \le r \le R} \Big\| \mathop{\mathbb{E}}_{r \le n \le r+L(r)} T^{\lfloor n + \log^3 n \rfloor} f \cdot T^n g_1 \cdot T^{\lfloor \log^2 n \rfloor} g_2 \Big\|_{L^2(\mu)}^2 = 0 \tag{3.9}$$

for some sub-linear function $L(t) \in \mathcal{H}$. In this example, we will choose the function $L(t)$, so that

$$\max_{r \le n \le r+L(r)} |\log^3(n) - \log^3(r)| = o_r(1) \quad \text{and} \quad \max_{r \le n \le r+L(r)} |\log^2(n) - \log^2(r)| = o_r(1).$$

For instance, the function $L(t) = \sqrt{t}$ can easily be checked to satisfy the above. Therefore, if r is very large, we can write

$$\Big\| \mathop{\mathbb{E}}_{r \le n \le r+L(r)} T^{\lfloor n + \log^3 n \rfloor} f \cdot T^n g_1 \cdot T^{\lfloor \log^2 n \rfloor} g_2 \Big\|_{L^2(\mu)} =$$
$$\Big\| \mathop{\mathbb{E}}_{r \le n \le r+L(r)} T^{n + \lfloor \log^3 r \rfloor + e_{1,n}} f \cdot\ T^n g_1 \cdot\ T^{\lfloor \log^2 r \rfloor + e_{2,n}} g_2 \Big\|_{L^2(\mu)},$$

where $e_{1,n}, e_{2,n} \in \{0, \pm 1\}$. We assume here that all the error terms are zero (in the main proof, we will invoke Lemma 3.2.1 to remove the error terms). Therefore, we want to show that

$$\limsup_{R\to+\infty} \mathop{\mathbb{E}}_{1\le r\le R} \Big\| \mathop{\mathbb{E}}_{r\le n\le r+L(r)} T^{n+\lfloor \log^3 r\rfloor} f \cdot\ T^n g_1 \cdot\ T^{\lfloor \log^2 r\rfloor} g_2 \Big\|_{L^2(\mu)}^2 = 0.$$

Since $\|g_2\|_\infty \le 1$, we reduce our problem to

$$\limsup_{R\to+\infty} \mathop{\mathbb{E}}_{1\le r\le R} \Big\| \mathop{\mathbb{E}}_{r\le n\le r+L(r)} T^n (g_1 \cdot T^{\lfloor \log^3 r\rfloor} f) \Big\|_{L^2(\mu)}^2 = 0.$$

Note that the inner average is a polynomial average in the variable n. We fix a positive integer M and use the van der Corput inequality to deduce that

$$\Big\| \mathop{\mathbb{E}}_{r\le n\le r+L(r)} T^n (g_1 \cdot T^{\lfloor \log^3 r\rfloor} f) \Big\|_{L^2(\mu)}^2 \ll \frac{1}{M} + \mathop{\mathbb{E}}_{|m|\le M} \Big| \int \overline{(g_1 \cdot T^{\lfloor \log^3 (r)\rfloor})} \cdot T^m (g_1 \cdot T^{\lfloor \log^3 (r)\rfloor})\ d\mu \Big| + o_r(1),$$

where the implied constant is absolute. Thus, we want to show that

$$\frac{1}{M} + \mathop{\mathbb{E}}_{|m|\le M} \mathop{\mathbb{E}}_{1\le r\le R} \Big| \int (\overline{g_1} \cdot T^m g_1) \cdot T^{\lfloor \log^3 (r)\rfloor} (\overline{f} \cdot T^m f)\ d\mu \Big| + o_R(1)$$

goes to 0, as $R \to +\infty$ and then as $M \to +\infty$.

Step 2: Dealing with the sub-linear functions.

Our problem follows by taking the limit as $R \to +\infty$ and then using the bound

$$\limsup_{R\to+\infty} \mathop{\mathbb{E}}_{1\le r\le R} \Big| \int (\overline{g_1} \cdot T^m g_1) \cdot T^{\lfloor \log^3 (r)\rfloor} (\overline{f} \cdot T^m f)\ d\mu \Big| \ll \|\!|\overline{f} \cdot T^m f|\!\|_{3,T}. \tag{3.10}$$

This was established in the previous example. Using this relation and taking the limit $M \to +\infty$ (note that our asymptotic constants do not depend on M), we reach the conclusion.

Since (3.10) follows from the previous example, we will describe our arguments for a more representative case. We shall prove that

$$\limsup_{N\to+\infty} \Big\| \mathop{\mathbb{E}}_{1\le n\le N} T^{\lfloor \log^3 n + \log^2 n\rfloor} f \cdot T^{\lfloor \log^3 n\rfloor} g_1 \cdot T^{\lfloor \log^2 n\rfloor} g_2 \Big\|_{L^2(\mu)} \ll \|\!|f|\!\|_4, \tag{3.11}$$

where the implied constant is absolute. Using Lemma 3.2.2, it suffices to show that

$$\limsup_{R\to+\infty} \mathop{\mathbb{E}}_{1\le r\le R} \Big\| \mathop{\mathbb{E}}_{r\le n\le r+L(r)} T^{\lfloor \log^3 n + \log^2 n\rfloor} f \cdot T^{\lfloor \log^3 n\rfloor} g_1 \cdot T^{\lfloor \log^2 n\rfloor} g_2 \Big\|_{L^2(\mu)}^2 \ll \|\!|f|\!\|_4^2$$

for some sub-linear function $L(t) \in \mathcal{H}$. We choose $L(t) = t(\log t)^{-3/2}$. Using similar approximations as in the first example, we can show that for any $0 \le h \le L(r)$

$$\log^3(r+h) = \log^3 r + h\frac{3\log^2 r}{r} + o_r(1),$$

while

$$\log^2(r+h) = \log^2 r + o_r(1)$$

for all $0 \le h \le L(r)$. Disregarding the error terms $o_r(1)$ in this example, it suffices to show that

$$\limsup_{R\to+\infty} \mathop{\mathbb{E}}_{1\le r\le R} \Big\| \mathop{\mathbb{E}}_{0\le h\le L(r)} T^{\left\lfloor \log^3 r + h\frac{3\log^2 r}{r}\right\rfloor} (T^{\lfloor \log^2 r\rfloor} f \cdot g_1) \cdot T^{\lfloor \log^2 r\rfloor} g_2 \Big\|_{L^2(\mu)}^2 \ll \|\!|f|\!\|_4^2.$$

Since g_2 is bounded by 1, the above bound follows from

$$\limsup_{N\to+\infty} \mathop{\mathbb{E}}_{1\le r\le R} \Big\| \mathop{\mathbb{E}}_{0\le h\le L(r)} T^{\left\lfloor \log^3 r + h\frac{3\log^2 r}{r}\right\rfloor} (T^{\lfloor \log^2 r\rfloor} f \cdot g_1) \Big\|_{L^2(\mu)} \ll \|\!|f|\!\|_4^2.$$

This is an average where the iterates are variable polynomials. Working similarly to the previous example, we can show that

$$\Big\|\mathop{\mathbb{E}}_{0\leq h\leq L(r)} T^{\left\lfloor \log^3 r+h\frac{3\log^2 r}{r}\right\rfloor}(T^{\lfloor \log^2 r\rfloor}f\cdot g_1)\Big\|^2_{L^2(\mu)} \ll$$
$$\frac{1}{M}+\mathop{\mathbb{E}}_{|m|\leq M}\Big|\int \overline{(T^{\lfloor \log^2 r\rfloor}f\cdot g_1)}\cdot T^m(T^{\lfloor \log^2 r\rfloor}f\cdot g_1)\, d\mu\Big| + o_r(1).$$

Thus, it suffices to show that

$$\limsup_{M\to+\infty}\mathop{\mathbb{E}}_{|m|\leq M}\limsup_{R\to+\infty}\mathop{\mathbb{E}}_{1\leq r\leq R}\Big|\int(\bar{g_1}\cdot T^m g_1)\cdot T^{\lfloor \log^2 r\rfloor}(\bar{f}\cdot T^m f)\, d\mu\Big| \ll \|\!|f|\!\|_4^2.$$

Note that we started with three sub-linear functions in the iterates and now we have an average with only one sub-linear function (our argument in the general case is based on this induction scheme). The result follows by working similarly to step 3 in the previous example.

3.4 Bounds for polynomial averages

Our main goal in this section is to establish Proposition 3.4.5 below. This will be based on the PET induction scheme, first introduced by Bergelson [1]. Before stating that proposition, we give some definitions.

3.4.1 Families of variable polynomials

Assume we are given a family $P_N = \{p_{1,N}, ..., p_{k,N}\}$ of essentially distinct (i.e. their pairwise differences are non-constant polynomials) variable polynomials, such that the degrees of the polynomials in P_N and of their pairwise differences are independent of N (for N large enough). Then, we assign to $p_{1,N}$ the vector $(v_{1,N}, ..., v_{k,N})$, where $v_{1,N}$ is the leading coefficient of $p_{1,N}$ and $v_{j,N}$ is the leading coefficient of $p_{1,N} - p_{j,N}$ for $j \neq 1$. We denote this by $\mathcal{S}(p_{1,N})$ and call this the *leading vector* of the family P_N corresponding to $p_{1,N}$. We similarly define $\mathcal{S}(p_{i,N})$ for every $i \in \{1, ..., k\}$ and call it the *leading vector* corresponding to $p_{i,N}$. Let us remark that the leading vector has no elements equal to 0, because we have assumed that the polynomials are essentially distinct. Finally, we call P_N *ordered*, if the degrees of the polynomials $p_{i,N}$ are non-increasing. In this case, the polynomial $p_{1,N}$ has maximal degree and we call it the *leading polynomial*. The *leading vector* of an ordered polynomial family is defined as the leading vector corresponding to its leading polynomial.

3.4.2 Types of polynomial families

We define the *type* $(d, w_d, ..., w_1)$ of the polynomial family, where d is the largest degree appearing in the polynomials of P_N and w_i is the number of distinct leading coefficients of the members of P_N with degree exactly i among all polynomials in the family. Note that for families of variable polynomials, the value of this vector may depend on the variable N. We order the types by the value of d and then order types of same degree lexicographically. We observe that a decreasing sequence of types must eventually be constant. The type of a family is a classical quantity used in the literature when an induction scheme on polynomial families is required.

3.4.3 Good sequences and nice polynomial families

Now, we define the notion of a nice polynomial family. Namely, we will deal with polynomials whose coefficients are well-behaved sequences. Our arguments fail to work in the general case where the coefficients can be arbitrary sequences.

Definition 3.4.1. *a) A sequence $(a_n)_{n\in\mathbb{N}}$ of real numbers is called "good", if there exists a function $f \in \mathcal{H}$ with $\lim_{t\to+\infty} f(t) \neq 0$ such that*

$$\lim_{n\to+\infty}\frac{a_n}{f(n)} = 1.$$

b) Let $P_N = \{p_{1,N}, ..., p_{k,N}\}$ be a collection of polynomials. The family P_N is called nice, if all the degrees of the polynomials $p_{i,N}$ and $p_{i,N} - p_{j,N}$ are independent of N for N large enough and their leading coefficients are good sequences, for all admissible values of the i, j.

Note that any good sequence has a limit (possibly infinite). An example of a good sequence that is not a Hardy sequence is the sequence $\frac{\lfloor N^{2/3} \rfloor}{\sqrt{N}}$, which behaves asymptotically like $N^{1/6}$. In general, all sequences of the form $\lfloor f(n) \rfloor$, where the function $f \in \mathcal{H}$ does not converge to 0 (as $t \to +\infty$), are good sequences, while, for example, $\left\lfloor \frac{1}{\log n} \right\rfloor$ is not a good sequence.

Lemma 3.4.2. *The type of a nice polynomial family is well-defined (independent of N) for N large enough.*

Proof. Assume that two polynomials $p_{i,N}$ and $p_{j,N}$ of the given family have the same degree s. Let $a_i(N), a_j(N), a_{ij}(N)$ be the leading coefficients of $p_{i,N}, p_{j,N}$ and $p_{i,N} - p_{j,N}$, which are all good sequences. The degree of the polynomial $p_{i,N} - p_{j,N}$ does not depend on N, for N sufficiently large, by the definition of a nice family. Then, we have either one of the following:
i) If the polynomial $p_{i,N} - p_{j,N}$ has degree equal to s, then for N large enough, $a_{ij}(N) = a_i(N) - a_j(N) \neq 0$ and therefore the polynomials $p_{i,N}, p_{j,N}$ have distinct leading coefficients eventually.
ii) If the polynomial $p_{i,N} - p_{j,N}$ has degree smaller than s, then that means that, for N large enough, we have $a_i(N) - a_j(N) = 0$ and the polynomials $p_{i,N}, p_{j,N}$ have equal leading coefficients eventually. The claim easily follows. □

3.4.4 The van der Corput inequality

We shall rely heavily on the following variant of the van der Corput inequality in our proofs.

Lemma 3.4.3. *For a sequence u_n in a Hilbert space with $\|u_n\| \leq 1$ and a quantity $M = o(N)$, we have*

$$\Big\|\frac{1}{N}\sum_{n=0}^{N-1} u_n\Big\|^{2^d} \ll_d \frac{1}{M} + \mathop{\mathbb{E}}_{-M\leq m\leq M} \Big|\mathop{\mathbb{E}}_{0\leq n\leq N-1} \langle u_{n+m}, u_n\rangle\Big|^{2^{d-1}} + o_N(1).$$

Proof. This follows from the basic van der Corput inequality

$$\Big\|\frac{1}{N}\sum_{n=0}^{N-1} u_n\Big\| \ll \frac{1}{M^{1/2}} + \Big(\mathop{\mathbb{E}}_{-M\leq m\leq M} \Big|\mathop{\mathbb{E}}_{0\leq n\leq N-1} \langle u_{n+m}, u_n\rangle\Big|\Big)^{1/2} + \frac{M^{1/2}}{N^{1/2}}$$

by successively squaring and applying the Cauchy-Schwarz inequality. □

We will use this inequality to derive asymptotic bounds for multiple ergodic averages involving polynomials. The above inequality holds, in particular, when M is a fixed positive integer. We state here the equivalent result for variable sequences, since this is more consistent with the notation used in the proof below.

Lemma 3.4.4. *For sequences $(u_{n,N})_{n,N\in\mathbb{N}}$ in a Hilbert space with $\|u_{n,N}\| \leq 1$ and a quantity $M = o(N)$, we have*

$$\Big\|\frac{1}{N}\sum_{n=0}^{N-1} u_{n,N}\Big\|^{2^d} \ll_d \frac{1}{M} + \mathop{\mathbb{E}}_{|m|\leq M} \Big|\mathop{\mathbb{E}}_{0\leq n\leq N-1} \langle u_{n+m,N}, u_{n,N}\rangle\Big|^{2^{d-1}} + o_N(1).$$

3.4.5 Bounds of polynomial averages

The remainder of the section will be dedicated to establishing the following proposition:

Proposition 3.4.5. *Let k, d be positive integers and let M be a positive integer parameter. Suppose $\mathbf{W} = (d, w_d, ..., w_1)$ is a (d+1)-tuple of positive integers that is also a type for some polynomial family. Then, there exist positive integers $t = t(d, k, \mathbf{W})$, $s = s(d, k, \mathbf{W})$, a finite set $Y = Y(d, k, \mathbf{W})$ of integers and integer polynomials in t variables $p_{\underline{\varepsilon},j}$, with $\underline{\varepsilon} \in [[s]]$ and $1 \leq j \leq k$, that are at most linear in each variable[3], such that for any ordered nice family of non-constant, essentially distinct polynomials*

$$P_N = \{p_{1,N}, ..., p_{k,N}\}$$

of degree d and type $\mathbf{W}$ with leading vector $\mathcal{S}(P_N) = \{u_{1,N}, ..., u_{k,N}\}$, any increasing sequence $L_N \to \infty$, any measure preserving system (X, μ, T) and sequences of 1-bounded functions $f_{1,N}, ..., f_{k,N}$, we have

$$\sup_{|c_{n,N}|\leq 1} \Big\| \mathop{\mathbb{E}}_{0\leq n\leq L_N} c_{n,N} \prod_{i=1}^{k} T^{\lfloor p_{i,N}(n) \rfloor} f_{i,N} \Big\|_{L^2(\mu)}^{2^t} \ll_{d,k,\mathbf{W}}$$
$$\frac{1}{M} + \sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} T^{\lfloor A_{\underline{\varepsilon},N}(\mathbf{m}) \rfloor + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} f_{1,N}) \, d\mu \Big| + o_N(1), \quad (3.12)$$

where

$$A_{\underline{\varepsilon},N}(\mathbf{m}) = \sum_{1\leq j\leq k} p_{\underline{\varepsilon},j}(\mathbf{m}) u_{j,N}$$

are real polynomials in $\mathbf{m}$. In addition, we have the following:
i) For $\underline{\varepsilon} \neq \underline{0}$, we have that the polynomial $A_{\underline{\varepsilon},N}(\mathbf{m})$ is non-constant.
ii) The polynomials $A_{\underline{\varepsilon},N}(\mathbf{m})$, $\underline{\varepsilon} \in [[s]]$ are pairwise essentially distinct.
iii) We have the relation

$$A_{\underline{\varepsilon},N}(\mathbf{m}) + A_{\underline{\varepsilon}^c,N}(\mathbf{m}) = A_{\underline{1},N}(\mathbf{m})$$

for any $\underline{\varepsilon} \in [[s]]$. More generally, if $\underline{\varepsilon}, \underline{\varepsilon}' \in [[s]]$ are such that $\underline{\varepsilon} + \underline{\varepsilon}' \in [[s]]$ [4], then

$$A_{\underline{\varepsilon},N}(\mathbf{m}) + A_{\underline{\varepsilon}',N}(\mathbf{m}) = A_{\underline{\varepsilon}+\underline{\varepsilon}',N}(\mathbf{m})$$

iv) For any $\underline{\varepsilon} \in [[s]]$,if

$$c_1 p_{\underline{\varepsilon},1}(\mathbf{m}) + ... + c_k p_{\underline{\varepsilon},k}(\mathbf{m})$$

is the zero polynomial for some $c_1, ..., c_k \in \mathbb{R}$, then we have $c_i = 0$ or $p_{\underline{\varepsilon},i}(\mathbf{m})$ is the zero polynomial, for every $1 \leq i \leq k$.

Remark. The $\sum_{\mathbf{h}\in Y^{[[s]]}}$ means that we take the sum for all choices of $\mathbf{h} = (h_{\underline{\varepsilon}},\ \underline{\varepsilon} \in [[s]])$ where $h_{\underline{\varepsilon}} \in Y$. In addition, we will make a small abuse of notation and write $\mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t}$ to denote the average over all $\mathbf{m} \in \mathbb{Z}^t \cap [-M, M]^t$.

Remarks. i) The polynomials $p_{\underline{\varepsilon},j}$ are independent of the leading vector $\{u_{1,N}, ..., u_{k,N}\}$ and are, more importantly, independent of the variable N.
ii) The existence of the errors $h_{\underline{\varepsilon}}$ is merely technical and arises from the floor function in the last expression inside the integral, since we cannot use Lemma 3.2.1 to remove the error terms in this case. This will appear in the proof of the case of linear polynomials that follows.
iii) The quantity $o_N(1)$ depends of course on the values of d and k. It also depends on the value of the fixed number M. However, this dependence plays no role in arguments of the following sections (where we will usually take limits first as $N \to +\infty$ and, then, as $M \to +\infty$). For ease of notation, we will omit all other subscripts for the term $o_N(1)$.
iv) The final condition *iv*) above implies that given a fixed $\underline{\varepsilon} \in [[s]]$, if we exclude all the constant polynomials among the $p_{\underline{\varepsilon},j}$, the remaining polynomials are linearly independent.

[3] This means that when regarded as polynomials only in one variable, then they are linear. Examples are $p_1(m_1, m_2) = m_1 - 2m_2$ and $p_2(m_1, m_2, m_3) = m_1 m_2 - 3m_3$.
[4] This means that the i-th entries of $\underline{\varepsilon}, \underline{\varepsilon}'$ cannot simultaneously be 1, for any admissible value of i.

Ignoring the technical parts of the statement, the above proposition asserts that when working with multiple averages on some polynomials that vary with N, we can instead bound them by the averages of a polynomial correlation sequence of only the function $f_{1,N}$. Even though the new polynomials $A_{\underline{\varepsilon},N}$ have several variables, they only depend on the sequences $u_{1,N},...,u_{k,N}$. Thus, if these sequences have good limiting behavior, we can take the limits first as $N\to+\infty$ and then as $M\to+\infty$ to get some nice bounds for the original averages. For instance, in the case where we have a fixed function $f_{1,N}=f_1$ and the sequences $u_{i,N}$ converge to non-zero real numbers, the above statement can be used to prove that the lim sup of the ergodic averages in the left-hand side of (3.12) can be bounded by a power of $\|\!| f_1 |\!\|_s$ for some suitable positive integer s. This last assertion follows from minor modifications to the argument present in [38] to cover the case of real polynomials (instead of just integer polynomials).

Proof of Proposition 3.14 in the linear case. We establish our proposition in the case where all the polynomials have degree 1. More precisely, assume that $p_{i,N}(t)=a_{i,N}t+b_{i,N}$ where $a_{i,N},b_{i,N}\in\mathbb{R}$ so that the variables $a_{i,N}$ are (eventually) non-zero. The assumption that our polynomials are essentially distinct implies that the numbers $a_{i,N}$ and $a_{j,N}$ are distinct. The leading vector of P_N is the set

$$\{a_{1,N},a_{1,N}-a_{2,N},...,a_{1,N}-a_{k,N}\}$$

and these are good sequences by our assumptions.

We induct on k. For $k=1$, we apply the van der Corput inequality to get

$$\Big\|\mathop{\mathbb{E}}_{0\le n\le L_N} c_{n,N}T^{\lfloor a_{1,N}n+b_{1,N}\rfloor}f_{1,N}\Big\|^2_{L^2(\mu)}\ll \\ \frac{1}{M}+\mathop{\mathbb{E}}_{|m|\le M}\Big|\mathop{\mathbb{E}}_{0\le n\le L_N}\overline{c_{n,N}}c_{n+m,N}\int\overline{f_{1,N}}\cdot T^{\lfloor a_{1,N}n+b_{1,N}+ma_{1,N}\rfloor-\lfloor a_{1,N}n+b_{1,N}\rfloor}f_{1,N}\,d\mu\Big|+o_N(1).$$

We rewrite the last quantity as

$$\frac{1}{M}+\mathop{\mathbb{E}}_{|m|\le M}\Big|\mathop{\mathbb{E}}_{0\le n\le L_N}\overline{c_{n,N}}c_{n+m,N}\int\overline{f_{1,N}}\cdot T^{\lfloor ma_{1,N}\rfloor+e_{n,m,N}}f_{1,N}\,d\mu\Big|+o_N(1),$$

where $e_{n,m,N}\in\{0,\pm1\}$ (the implied constant is independent of all variables in the above relation). Let $A_{z,m,N}=\{n\in\mathbb{Z}^+\colon 0\le n\le L_N \text{ and } e_{n,m,N}=z\}$ for $z\in\{0,\pm1\}=Y$. Then, the innermost average can be rewritten as

$$\Big|\frac{1}{L_N}\sum_{z\in Y}\sum_{n\in A_{z,N,m}}\overline{c_{n,N}}c_{n+m,N}\int\overline{f_{1,N}}\cdot T^{\lfloor ma_{1,N}\rfloor+z}f_{1,N}\,d\mu\Big|\le\sum_{z\in Y}\Big|\int\overline{f_{1,N}}\cdot T^{\lfloor ma_{1,N}\rfloor+z}f_{1,N}\,d\mu\Big|,$$

which, combined with the above, gives the desired result (for constants $t=1$ and $s=1$, polynomials $p_1(m)=ma_{1,N}$ and $p_0(m)=0$ and set $Y=\{0,\pm1\}$).

Now assume that we have proven the result for $k-1$ ($k\ge2$), with the constants of the proposition given by $t=k-1$ and $s=k-1$. Then, we use the van der Corput inequality to get

$$\Big\|\mathop{\mathbb{E}}_{0\le n\le L_N} c_{n,N}\prod_{i=1}^{k}T^{\lfloor a_{i,N}n+b_{i,N}\rfloor}f_{i,N}\Big\|^{2^k}_{L^2(\mu)}\ll_k\frac{1}{M}+o_N(1)+ \\ \mathop{\mathbb{E}}_{|m|\le M}\Big|\mathop{\mathbb{E}}_{0\le n\le L_N}\overline{c_{n,N}}c_{n+m,N}\int\prod_{i=1}^{k}T^{\lfloor a_{i,N}n+b_{i,N}\rfloor+\lfloor ma_{i,N}\rfloor+e_{i,m,n,N}}f_{1,N}\,T^{\lfloor a_{i,N}n+b_{i,N}\rfloor}\overline{f_{i,N}}\,d\mu\Big|^{2^{k-1}},$$

which is smaller than

$$\mathop{\mathbb{E}}_{|m|\le M}\Big|\mathop{\mathbb{E}}_{0\le n\le L_N}\overline{c_{n,N}}c_{n+m,N}\int\prod_{i=1}^{k}T^{\lfloor a_{i,N}n+b_{i,N}\rfloor-\lfloor a_{k,N}n+b_{k,N}\rfloor+\lfloor ma_{i,N}\rfloor+e_{i,m,n,N}}f_{1,N}\cdot \\ T^{\lfloor a_{i,N}n+b_{i,N}\rfloor-\lfloor a_{k,N}n+b_{k,N}\rfloor}\overline{f_{i,N}}\,d\mu\Big|^{2^{k-1}}+1/M+o_N(1),\quad(3.13)$$

where we again have $e_{i,m,,n,N} \in \{0, \pm 1\}$. In the last step, we composed with $T^{-\lfloor a_{k,N}n+b_{k,N}\rfloor}$ inside the integral.

We have

$$\lfloor a_{i,N}n + b_{i,N}\rfloor - \lfloor a_{k,N}n + b_{k,N}\rfloor = \lfloor (a_{i,N} - a_{k,N})n + b_{i,N} - b_{k,N}\rfloor + e'_{i,n,N}$$

where $e'_{i,n,N} \in \{0, \pm 1\}$. Therefore, we can rewrite the last expression in (3.13) as

$$\frac{1}{M} + \underset{|m|\leq M}{\mathbb{E}} \Big| \underset{0\leq n\leq L_N}{\mathbb{E}} \overline{c_{n,N}} c_{n+m,N} \prod_{i=1}^{k} \int T^{\lfloor (a_{i,N}-a_{k,N})n+b_{i,N}-b_{k,N}\rfloor + e'_{i,n,N}} \big(\overline{f_{i,N}} \cdot T^{\lfloor ma_{i,N}\rfloor + e_{i,m,n,N}} f_{i,N}\big)\, d\mu \Big|^{2^{k-1}} + o_N(1).$$

Then, using the Cauchy-Schwarz inequality and the argument in Lemma 3.2.1, we can bound the innermost average in the above expression by $O_k(1)$ times the quantity

$$A_{m,N} = \sup_{|c_{n,N}|\leq 1} \Big\| \underset{0\leq n\leq L_N}{\mathbb{E}} c_{n,N} \prod_{i=1}^{k-1} T^{\lfloor (a_{i,N}-a_{k,N})n + (b_{i,N}-b_{k,N})\rfloor} \big(\overline{f_{i,N}} \cdot T^{\lfloor ma_{i,N}\rfloor + e_{i,m,n,N}} f_{i,N}\big) \Big\|_{L^2(\mu)}^{2^{k-1}}.$$

Now, we use the argument of Lemma 3.2.1 again to deduce that $A_{m,N}$ is bounded by $O_k(1)$ times

$$\sum_{\substack{z_i \in \{0,\pm 1\} \\ 1\leq i\leq k-1}} \sup_{|c_{n,N}|\leq 1} \Big\| \underset{0\leq n\leq L_N}{\mathbb{E}} c_{n,N} \prod_{i=1}^{k-1} T^{\lfloor (a_{i,N}-a_{k,N})n + (b_{i,N}-b_{k,N})\rfloor} \big(\overline{f_{i,N}} \cdot T^{\lfloor ma_{i,N}\rfloor + z_i} f_{i,N}\big) \Big\|_{L^2(\mu)}^{2^{k-1}}.$$

We fix some $\mathbf{z} = (z_1, ..., z_{k-1}) \in \{0, \pm 1\}^{k-1}$. If we take the polynomial that corresponds to $\overline{f_{1,N}} \cdot T^{\lfloor ma_{1,N}\rfloor + z_1} f_{1,N}$ to be the new leading polynomial, then the new leading vector is the set

$$\{a_{1,N} - a_{k,N}, a_{1,N} - a_{2,N}, ..., a_{1,N} - a_{k-1,N}\}.$$

By the induction hypothesis, there exists a finite set Y_{k-1}, for which

$$\sup_{|c_{n,N}|\leq 1} \Big\| \underset{0\leq n\leq L_N}{\mathbb{E}} c_{n,N} \prod_{i=1}^{k-1} T^{\lfloor (a_{i,N}-a_{k,N})n + (b_{i,N}-b_{k,N})\rfloor} \big(\overline{f_{i,N}} \cdot T^{\lfloor ma_{i,N}\rfloor + z_i} f_{i,N}\big) \Big\|_{L^2(\mu)}^{2^{k-1}} \ll_k$$
$$\frac{1}{M} + \sum_{\mathbf{h}\in[[Y_{k-1}]]} \underset{|m_1|,...,|m_{k-1}|\leq M}{\mathbb{E}} \Big| \int \prod_{\underline{\varepsilon}\in[[k-1]]} T^{\lfloor \sum_{1\leq j\leq k-1} p_{\underline{\varepsilon},j}(m_1,...,m_{k-1})(a_{1,N}-a_{j,N})\rfloor + h_{\underline{\varepsilon}}}$$
$$\mathcal{C}^{|\underline{\varepsilon}|} \big(\overline{f_{1,N}} \cdot T^{\lfloor ma_{1,N}\rfloor + z_1} f_{1,N}\big)\, d\mu \Big| + o_N(1).$$

Using the identification $[[k]] = [[k-1]] \times \{0, 1\}$, we can write an $\underline{\varepsilon} \in [[k]]$ as $\underline{\varepsilon} = (\underline{\varepsilon_1}, \varepsilon_2)$ where $\underline{\varepsilon_1} \in [[k-1]]$ and $\varepsilon_2 \in \{0, 1\}$. We also write $\mathbf{m} = (m, m_1, ..., m_{k-1})$. Combining the integer parts, we rewrite the last integral as

$$\int \prod_{\underline{\varepsilon}\in[[k]]} T^{\lfloor \sum_{1\leq j\leq k-1} p'_{\underline{\varepsilon},j}(m_1,...,m_{k-1})(a_{1,N}-a_{j,N}) + p'_{\underline{\varepsilon},k}(m)a_{1,N}\rfloor + h'_{\underline{\varepsilon},\mathbf{m}}} \mathcal{C}^{|\underline{\varepsilon}|} f_{1,N}\, d\mu\ ,$$

where

1. $p'_{\underline{\varepsilon},j}$ is the polynomial $p_{\underline{\varepsilon_1},j}$ for $1 \leq j \leq k-1$,
2. the polynomial $p'_{\underline{\varepsilon},k}$ is equal to m when $\underline{\varepsilon_2} = 0$ and is zero otherwise and
3. $h'_{\underline{\varepsilon},\mathbf{m}} = h_{\underline{\varepsilon_1}} + h_{2,\underline{\varepsilon},\mathbf{m}}$, where[5] $h_{2,\underline{\varepsilon},\mathbf{m}} \in \{0, \pm 1, \pm 2\}$. More importantly, $h'_{\underline{\varepsilon},m}$ takes values in a finite set Y_k.

[5]In particular, $h_{2,\underline{\varepsilon},\mathbf{m}}$ is the sum of z_1 plus the error term appearing by combining $\lfloor ma_{1,N}\rfloor$ with the other integer part, whenever they both appear. Otherwise, it is zero. Thus, it takes values on a finite set of integers.

We observe that

$$\Big|\int \prod_{\underline{\varepsilon}\in[[k]]} T^{\left\lfloor \sum_{1\leq j\leq k-1} p'_{\underline{\varepsilon},j}(m_1,\dots,m_{k-1})(a_{1,N}-a_{j,N})+p'_{\underline{\varepsilon},k}(m)a_{1,N}\right\rfloor + h'_{\underline{\varepsilon},\mathbf{m}}} \mathcal{C}^{|\underline{\varepsilon}|} f_{1,N}\, d\mu\Big| \leq$$

$$\sum_{\mathbf{h}\in[[Y_k]]} \Big|\int \prod_{\underline{\varepsilon}\in[[k]]} T^{\left\lfloor \sum_{1\leq j\leq k-1} p'_{\underline{\varepsilon},j}(m_1,\dots,m_{k-1})(a_{1,N}-a_{j,N})+p'_{\underline{\varepsilon},k}(m)a_{1,N}\right\rfloor + h_{\underline{\varepsilon}}} \mathcal{C}^{|\underline{\varepsilon}|} f_{1,N}\, d\mu\Big|.$$

Averaging over $m, m_1, ..., m_{k-1}$ and summing over $\mathbf{z} \in \{0, \pm 1\}^{k-1}$, we have that for the finite set Y_k above, the original expression is bounded by $O_k(1)$ times

$$\frac{1}{M} + \sum_{\mathbf{h}\in[[Y_k]]} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^k} \Big|\int \prod_{\underline{\varepsilon}\in[[k]]} T^{\left\lfloor \sum_{1\leq j\leq k} p'_{\underline{\varepsilon},j}(\mathbf{m})u_{j,N}\right\rfloor + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} f_{1,N})\, d\mu\Big| + o_N(1),$$

where $u_{1,N} = a_{1,N}$ and $u_{j,N} = a_{1,N} - a_{j,N}$. The conclusion follows. □

Remark. It follows from the above proof that the polynomials $A_{\underline{\varepsilon},N}$ in the statement of Proposition 3.4.5 have the following form:

$$A_{\underline{\varepsilon},N}(m_1, ..., m_k) = \underline{\varepsilon} \cdot (m_1u_{1,N}, ..., m_ku_{k,N})$$

where "$\cdot$" denotes here the standard inner product on $\mathbb{R}^k$. Thus, it is straightforward to check that the polynomials $A_{\underline{\varepsilon},N}$ satisfy the conditions $i), ii), iii)$ and $iv)$ of Proposition 3.4.5. Note that all these polynomials have degree 1. This will not be the case when working with polynomials of higher degree, where we may have higher degree terms (like products of the form m_1m_2), but they will be linear in each variable separately.

3.4.6 The PET induction.

For a polynomial p_N, a family P_N and $h \in \mathbb{N}$, we define the *van der Corput* operation (or vdC operation), by replacing our original family with the family

$$\{p_{1,N}(t+h) - p_N(t), ..., p_{k,N}(t+h) - p_N(t),\ p_{1,N}(t) - p_N(t), ..., p_{k,N}(t) - p_N(t)\}$$

and then removing polynomials of degree 0. We denote this new family by $(p_N, h)^*P_N$. At first glance, it is not obvious that this operation is well defined, because the constant polynomials that we discard may be different for different values of N. We will see that this is not the case for nice polynomial families below. We will use the vdC operation successively to reduce the "complexity" of a polynomial family. Our main observation is that the leading vector of a polynomial family is well behaved under the vdC operation.

Consider a family of variable polynomials $P_N = \{p_{1,N}, ..., p_{k,N}\}$ and let the leading vector of P_N corresponding to $p_{1,N}$ be

$$\mathcal{S}(P_N) = \{u_{1,N}, ..., u_{k,N}\}.$$

Fix any $1 \leq i_0 \leq k$, as well as the polynomial $p_{i_0,N}$, which we denote as p_N from now on for convenience. Consider the new polynomial family $P'_{N,h} = (p_N, h)^{\star}P_N$ that arises from the van der Corput operation. Here, h ranges over the non-zero integers.

Lemma 3.4.6. *Assume that the family P_N of degree d is nice and let $(u_{1,N}, ..., u_{k,N})$ be its leading vector corresponding to $p_{1,N}$. For every choice of polynomial p_N above and the value of $h \in \mathbb{Z}^*$, we have that each element of the leading vector of $P'_{N,h}$ corresponding to the new polynomial $p_{1,N}(t+h) - p_N(t)$ has one the following forms:*

- *They are equal to one of the $u_{i,N}$ for some $2 \leq i \leq k$.*
- *They have the form $du_{1,N}h$.*
- *They are the sum $du_{1,N}h + u_{i,N}$ for some $u_{i,N}$ with $i \neq 1$.*

Proof. Without loss of generality, we will assume that we have taken $p_N = p_{k,N}$ (the case $p_N = p_{1,N}$ is very similar). We want to study the leading vector corresponding to the polynomial $p_{1,N}(t+h)-p_{k,N}(t)$. Therefore, it is sufficient to find the leading coefficients of the polynomials

$$\big(p_{1,N}(t+h) - p_{k,N}(t)\big) - \big(p_{1,N}(t) - p_{k,N}(t)\big)$$
$$\big(p_{1,N}(t+h) - p_{k,N}(t)\big) - \big(p_{i,N}(t+h) - p_{k,N}(t)\big)$$
$$\big(p_{1,N}(t+h) - p_{k,N}(t)\big) - \big(p_{i,N}(t) - p_{k,N}(t)\big)$$

for $2 \leq i \leq k$. The leading coefficient of the first polynomial is always $dhu_{1,N}$ and that satisfies our required property. The leading coefficient of the second polynomial is always equal to the leading coefficient of $p_{1,N}(t+h)-p_{i,N}(t+h)$ and this is always equal to the leading coefficient of $p_{1,N}(t)-p_{i,N}(t)$ which belongs to the leading vector. Finally, the leading coefficient of the third polynomial is equal to the leading coefficient of $p_{1,N}(t + h) - p_{i,N}(t)$. Note that this polynomial can be rewritten as

$$(p_{1,N}(t + h) - p_{1,N}(t)) + (p_{1,N}(t) - p_{i,N}(t)).$$

The leading coefficient of the first polynomial is equal to $dhu_{1,N}$ as we established above, while the second difference has leading coefficient $u_{i,N}$ (by definition). Therefore, the leading coefficient of their sum is either $dhu_{1,N}, u_{i,N}$ or their sum $dhu_{1,N} + u_{i,N}$, which concludes the proof. □

Observe that the particular form each element of the new leading vector is independent of N (i.e. it cannot have the first form for one value of N and then the second form for some other value of N). This follows from the fact that the type of the original family is independent of N, if N is large enough. We will now use this lemma to study how the van der Corput operation affects the type of the original family.

Corollary 3.4.7. *Let P_N, p_N be as above and let d be the degree of the family P_N. Then, there exists a set of integers Y with at most $O_{k,d}(1)$ elements such that, for every $h \notin Y$, the polynomial family $P'_{N,h} = (p_N, h)^{\star} P_N$ that arises from the van der Corput operation is nice and its type is independent[6] of the value of h.*

Proof. We denote by $u_{ii,N}$ the leading coefficient of $p_{i,N}$, while $u_{ij,N}$ denotes the leading coefficient of $p_{i,N} - p_{j,N}$ for $i \neq j$. These are all good sequences by the definition of a nice family. Using Lemma 3.4.6, we can prove that the leading coefficients of all the polynomials in $P'_{N,h}$ and of their differences can take one of the following forms:
i) they are equal to some $u_{ij,N}$ with $i \neq j$,
ii) they have the form $ru_{ii,N}h$ for some $1 \leq r \leq d$ or
iii) they have the form $ru_{ii,N}h + u_{ij,N}$ for some $1 \leq r \leq d$.

We prove that these sequences are good for all except $O_{d,k}(1)$ values of h. For all values of $1 \leq i, j \leq k$ $(i \neq j)$ and $1 \leq r \leq d$, we consider the set $A(i, j, r)$ of all possible sequences of the above three forms (not all of them appear as leading coefficients, but this does not affect our argument), where h is some fixed non-zero integer. There are only finitely many such sets. Note that for $h \neq 0$, the sequences of the first two forms are always good. Now consider a sequence of the form $ru_{ii,N}h + u_{ij,N}$. There exist functions $f_1, f_2 \in \mathcal{H}$, not converging to 0, such that $|u_{ii,N}/f_1(N)| = 1 + o_N(1)$ and $|u_{ij,N}/f_2(N)| = 1 + o_N(1)$. The function $rhf_1(t) + f_2(t)$ is obviously an element of $\mathcal{H}$. In addition, for our fixed r, the relation

$$\lim_{t\to+\infty} (rhf_1(t) + f_2(t)) = 0$$

can hold only for at most one possible value of $h \in \mathbb{Z}$, which we call a "bad value". Then, if h is not a bad value, we have

$$\Big|\frac{ru_{ii,N}h + u_{ij,N}}{rhf_1(N) + f_2(N)}\Big| = 1 + o_N(1).$$

Indeed, this follows easily because the functions f_1 and f_2 are comparable, which also means that all the sequences involved are comparable. Thus, dividing the numerator and denominator of the above

[6] The type depends only on which polynomial of the initial family we choose to be the polynomial p_N, as well as the type of the original family.

fraction by either $f_1(N)$ or $f_2(N)$, we easily get the last assertion. In short, we have shown that the sequence $ru_{ii,N}h + u_{ij,N}$ is a good sequence for all non-bad values of h. Now, if we take all possible values of the i, j, r, we conclude that there are at most $O_{d,k}(1)$ bad values of h.

We have shown that for every non-bad value of h, the family $P'_{N,h}$ is a nice polynomial family and, therefore, has a fixed type (independent of N). We show that its type does not depend on h. In order to prove this, we consider two polynomials q_1, q_2 of $P'_{N,h}$ of the same degree. We consider some possible cases:
a) If q_1 and q_2 have the form $p_{i,N}(t) - p_N(t)$, then whether or not their leading coefficients are equal depends only on the type of the original family and the choice of p_N (and not on h).
b) If q_1 has the form $p_{i,N}(t+h) - p_N(t)$, while q_2 has the form $p_{j,N}(t) - p_N(t)$, then their leading coefficients can be equal in only two possible cases: if the polynomial p_N has degree strictly larger than the degree of both $p_{i,N}$ and $p_{j,N}$ (this depends only on the choice of p_N, not on h), or if the polynomials $p_{i,N}(t+h)$ and $p_{j,N}(t)$ have the same degree (bigger than or equal to the degree of p_N) and equal leading coefficients. In the second case, we must have that $p_{i,N}(t)$ and $p_{j,N}(t)$ have equal leading coefficients, which depends only on the type of the original family and not on h.
c) If q_1 and q_2 both have the form $p_{i,N}(t+h) - p_N(t)$, then the result follows similarly as in the case a).

The fact that the degrees of the polynomials of the new family and of their differences do not depend on N and h can also be established easily using the preceding arguments. We omit the details. □

Proposition 3.4.8. *If $P_N = \{p_{1,N}, ..., p_{k,N}\}$ is an ordered polynomial family, then there exists a polynomial $p_N \in P_N$, such that for all, except at most one value of $h \in \mathbb{Z}$, the polynomial family $P'_{N,h} = (p_N, h)^* P_N$ has type strictly smaller than the type of P_N and its leading polynomial is the polynomial $p_{1,N}(t+h) - p_N(t)$.*

Proof. We describe the operation that reduces the type. At each step, we choose a polynomial $p_N \in P_N$ that has minimal degree in the family. For an $h \in \mathbb{Z}$, apply the van der Corput operation. This forms a polynomial family

$$P'_N = \{p_{1,N}(t+h) - p_N(t), ..., p_{k,N}(t+h) - p_N(t),\ p_{1,N}(t) - p_N(t), ..., p_{k,N}(t) - p_N(t)\} \tag{3.14}$$

and choose $p_{1,N}(t+h) - p_N(t)$ to be the new leading polynomial. We distinguish between some cases:

a) Assume that the polynomials $p_{1,N}$ and $p_{k,N}$ have distinct degrees. Then, choose $p_N = p_{k,N}$, which by the "ordered" assumption has minimal degree. We notice that the polynomial $p_{1,N}(t+h) - p_N(t)$ has maximal degree in the polynomial family. We check that the type of the polynomial family is reduced. Indeed, if the degree of $p_{k,N}(t)$ is d', then the number w'_d is reduced, while all the numbers w_i are left unchanged for $i > d'$.

b) Suppose the polynomials $p_{1,N}$ and $p_{k,N}$ have the same degree and not all leading coefficients in the family P_N are equal. In particular, we may assume, without loss of generality, that this holds for the polynomials $p_{1,N}$ and $p_{k,N}$. Again, choose $p_N = p_{k,N}$. Then, the polynomial $p_{1,N}(t+h) - p_N(t)$ has maximal degree in the new polynomial family. In addition, the number w_d is reduced, which means that the new family has smaller type than the original.

c) Finally, assume that all polynomials have the same degree and the same leading coefficient. We choose again $p_N = p_{k,N}$. The polynomial $p_{1,N}(t+h) - p_N(t)$ has maximal degree equal to $d-1$ in P'_N, except possibly for one value of $h \in \mathbb{Z}$ (to see this, we can work similarly as in the proof of Corollary 3.4.7). Also, the family P'_N has smaller type than P_N, since it has degree at most $d-1$. □

While for a given type $\mathbf{W}$ there are infinite types smaller than $\mathbf{W}$, it is straightforward to see that a decreasing sequence of types is eventually constant. Therefore, the type-reducing operation that we did above will eventually terminate to a type of degree 1, namely we will reduce our problem to the linear case, which we have already established. To summarize all of the above, we have the following:

Corollary 3.4.9. *Let P_N be a nice polynomial family of degree d, with k polynomials and with type $\mathbf{W}$. Then, there exists a $p_N \in P_N$, such that the family $P'_N = (p_N, h)^* P_N$ is nice and has (fixed) type smaller than $\mathbf{W}$ for all, except at most $O_{d,k}(1)$ values of h.*

Definition 3.4.10. *We will call a van der Corput operation* $(p_N, h)^* P_N$ non-degenerate*, if the polynomial* $p_N \in P_N$ *is such, that the conditions of Corollary 3.4.9 hold.*

Namely, the polynomial p_N must be chosen, so that the resulting family has type independent of N, h, provided that N is sufficiently large and h takes values outside a set of at most $O_{d,k}(1)$ elements (here, this notation refers to the same asymptotic constant appearing in the statement of Corollary 3.4.9). In view of the above corollary, we deduce that there always exists a non-degenerate van der Corput operation. We will denote a non-degenerate van der Corput operation simply by $(p_N)^* P_N$ to indicate the independence on the parameter h.

We are now ready to finish the proof of Proposition 3.4.5:

Proof of the higher degree case. First of all, we shall explain how we will choose the parameters t, s. These depend crucially on how the van der Corput operations are used (and there are possibly many ways in which the successive van der Corput can be carried out), which may lead to ambiguity.

Let $\mathbf{W} = (d, w_d, ..., w_1)$ be the type of the given polynomial family. We say that a triplet $(d', k', \mathbf{W}')$ can be reached by the triplet $(d, k, \mathbf{W})$ if there exists a sequence of non-degenerate van der Corput operations that produces the families

$$P_{1,N} = (q_{1,N})^* P_N, ..., P_{\ell,N} = (q_{\ell,N})^* P_{\ell-1,N},$$

where the family $P_{\ell,N}$ consists of k' polynomials, has degree d' and type $\mathbf{W}'$.

Observe that the triplets that can be reached by the original triplet $(d, k, \mathbf{W})$ are finitely many in number, since there are only finitely many choices (depending on $d, k, \mathbf{W}$) for each polynomial $q_{i,N}$ at each step. In particular, they all have degrees at most d, types strictly smaller than $\mathbf{W}$ and the number k' can be bounded by a function of $(d, k, \mathbf{W})$, since each van der Corput operation at most doubles the number of polynomials in a family and this operation can occur finitely many times as well. We also remind the reader that we have already established our claim for all polynomial families of degree $d = 1$ and will serve as the base case of our induction).

Let $S_{d,k,\mathbf{W}}$ be the set of triplets that can be possibly reached by $(d, k, \mathbf{W})$, which is a finite set. We will use induction by considering that our claim holds for all triplets in $S_{d,k,\mathbf{W}}$ and we will show that the claim holds for families corresponding to our original family P_N that corresponds to the triplet $(d, k, \mathbf{W})$.

Fix such a triplet $(d', k', \mathbf{W}')$ and define $t(d', k', \mathbf{W}'), s(d', k', \mathbf{W}')$ to be the numbers appearing in the statement of Proposition 3.4.5. Namely, if the nice ordered family

$$Q_N = \{q_{1,N}, ..., q_{k',N}\}$$

has degree d' and type $\mathbf{W}'$, then

$$\sup_{|c_{n,N}| \leq 1} \Big\| \mathop{\mathbb{E}}_{0 \leq n \leq L_N} c_{n,N} \prod_{i=1}^{k'} T^{\lfloor q_{i,N}(n) \rfloor} f_{i,N} \Big\|_{L^2(\mu)}^{2^{t(d',k',\mathbf{W}')}} \ll_{d',k',\mathbf{W}'}$$
$$\frac{1}{M} + \sum_{\mathbf{h} \in Y^{[[s(d',k',\mathbf{W}')]]}} \mathop{\mathbb{E}}_{\mathbf{m} \in [-M,M]^{t(d',k',\mathbf{W}')}} \Big| \int \prod_{\underline{\varepsilon} \in [[s(d',k',\mathbf{W}')]]} T^{\lfloor A_{\underline{\varepsilon},N}(\mathbf{m}) \rfloor + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} f_{1,N}) \, d\mu \Big| + o_N(1), \quad (3.15)$$

where we are being vague on the dependence of the polynomials $A_{\underline{\varepsilon},N}$ on the parameters $(d', k', \mathbf{W}')$ and the family Q_N in this relation, since this will not concern us temporarily.

The number $t(d', k', \mathbf{W}')$ is the number of times we apply the van der Corput inequality in order to bound the left-hand side by the quantity on the right-hand side. Now, we define

$$t_0 = \max_{(d',k',\mathbf{W}') \in S_{d,k,\mathbf{W}}} t(d', k', \mathbf{W}'),$$

which, of course, is a parameter that depends only on $(d, k, \mathbf{W})$. Assume that the number t_0 corresponds to a family Q_N. Then, it is obvious that Q_N can be reached by the original family P_N in only one step. Indeed, if there was a another family in the sequence of van der Corput operations starting from P_N to Q_N, then this family would have a strictly larger parameter $t(\cdot)$ associated to it than t_0.

Assume that the family Q_N has the triplet $(d', k', \mathbf{W}')$ associated to it. Although the parameter t_0 is well defined (and depends only on $(d, k, \mathbf{W})$), the parameter $s(d', k', \mathbf{W}')$ may not be, because there may be another family Q'_N which has the same value t_0 for the first parameter, but different for the second. In this case, we simply take Q_N to be the one for which the parameter $s(d', k', \mathbf{W})$ is also maximized (denote this simply by s from this point onward). Obviously, we have that s depends only on $(d, k, \mathbf{W})$.

For the family Q_N constructed above, we can write $Q_N = (p_N)^* P_N$ for some $p_N \in P_N$. Without loss of generality, assume that $p_N = p_{k,N}$ (the case where $p_N = p_{1,N}$ is similar).

We apply the van der Corput inequality to get

$$\Big\|\mathop{\mathbb{E}}_{0\leq h\leq L_N} c_{n,N}\prod_{i=1}^{k} T^{\lfloor p_{i,N}(n)\rfloor} f_{i,N}\Big\|_{L^2(\mu)}^{2^{t_0+1}} \ll_{t_0}$$

$$\frac{1}{M} + \mathop{\mathbb{E}}_{|m|\leq M}\Big|\mathop{\mathbb{E}}_{0\leq n\leq L_N} c_{n+m,N}\overline{c_{n,N}}\int\prod_{i=1}^{k} T^{\lfloor p_{i,N}(n+m)\rfloor} f_{i,N}\cdot T^{\lfloor p_{i,N}(n)\rfloor}\overline{f_{i,N}}\,d\mu\Big|^{2^{t_0}} + o_N(1). \quad (3.16)$$

We compose with $T^{-\lfloor p_N(n)\rfloor}$ in the above integral, so that

$$\mathop{\mathbb{E}}_{0\leq n\leq L_N} c_{n+m,N}\overline{c_{n,N}}\int\prod_{i=1}^{k} T^{\lfloor p_{i,N}(n+m)\rfloor} f_{i,N}\cdot T^{\lfloor p_{i,N}(n)\rfloor}\overline{f_{i,N}}\,d\mu =$$

$$\mathop{\mathbb{E}}_{0\leq n\leq L_N} c_{n+m,N}\overline{c_{n,N}}\int\prod_{i=1}^{k} T^{\lfloor p_{i,N}(n+m)\rfloor-\lfloor p_N(n)\rfloor} f_{i,N}\cdot T^{\lfloor p_{i,N}(n)\rfloor-\lfloor p_N(n)\rfloor}\overline{f_{i,N}}\,d\mu =$$

$$\mathop{\mathbb{E}}_{0\leq n\leq L_N} c_{n+m,N}\overline{c_{n,N}}\int\prod_{i=1}^{k} T^{\lfloor p_{i,N}(n+m)-p_N(n)\rfloor+e_{1,n,i,m,N}} f_{i,N}\cdot T^{\lfloor p_{i,N}(n)-p_N(n)\rfloor+e_{2,n,i,N}}\overline{f_{i,N}}\,d\mu,$$

where the numbers $e_{1,n,i,m,N}$ and $e_{2,n,i,N}$ take values in the set $\{0, \pm 1\}$. We use the Cauchy-Schwarz inequality and then use Lemma 3.2.1 to bound the absolute value of the last quantity by a constant (depending only on k) multiple of the expression

$$\sup_{|c_{n,N}|\leq 1}\Big\|\mathop{\mathbb{E}}_{0\leq n\leq L_N} c_{n,N}\Big(\prod_{i=1}^{k-1} T^{\lfloor p_{i,N}(n+m)-p_N(n)\rfloor} f'_{i,N}\cdot T^{\lfloor p_{i,N}(n)-p_N(n)\rfloor}\overline{f'_{i,N}}\Big)$$

$$T^{\lfloor p_{k,N}(n+m)-p_{k,N}(n)\rfloor} f'_{k,N}\Big\|_{L^2(\mu)} + o_N(1)$$

for some 1-bounded functions $f'_{1,N} = f_{1,N}, f'_{2,N}, ..., f'_{k,N}$. Recall that we chose $p_N = p_{k,N}$. The family of polynomials

$$P'_{N,m} = \{p_{1,N}(t+m) - p_{k,N}(t), ..., p_{k,N}(t+m) - p_{k,N}(t),\ p_{1,N}(t) - p_{k,N}(t), ..., p_{k-1,N}(t) - p_{k,N}(t)\}$$

is nice and has (fixed) type $\mathbf{W}' < \mathbf{W}$ independent of m for all, except at most $O_{d,k}(1)$ values of $m \in \mathbb{N}$ (it has the same triplet $(d', k', \mathbf{W}')$ of parameters as the family Q_N above). Let Q be this finite set of "bad" values of m and let

$$\mathcal{S}(P'_{N,m}) = \{u'_{1,m,N}, ..., u'_{k',m,N}\}$$

be the leading vector of $P'_{N,m}$, where $k' \leq 2k-1$. For all $m \notin Q$, we use the induction hypothesis to deduce that

$$\sup_{|c_{n,N}|\leq 1}\Big\|\mathop{\mathbb{E}}_{0\leq n\leq L_N} c_{n,N}\Big(\prod_{i=1}^{k-1} T^{\lfloor p_{i,N}(n+m)-p_N(n)\rfloor} f'_{i,N}\cdot T^{\lfloor p_{i,N}(n)-p_N(n)\rfloor}\overline{f'_{i,N}}\Big)\cdot$$

$$T^{\lfloor p_{k,N}(t+m)-p_{k,N}(t)\rfloor} f'_{k,N}\Big\|_{L^2(\mu)}^{2^{t_0}} \ll_{k,d,\mathbf{W}'}$$

$$\frac{1}{M} + \sum_{\mathbf{h}\in[[Y_0]]}\mathop{\mathbb{E}}_{(m_1,...,m_t)\in[-M,M]^t}\Big|\int\prod_{\underline{\varepsilon}\in[[s]]} T^{\lfloor\sum_{1\leq j\leq k'} p_{\underline{\varepsilon},j}(m_1,...,m_t)u'_{j,m,N}\rfloor+h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}f_{1,N})\,d\mu\Big| + o_N(1) \quad (3.17)$$

for a finite set Y_0 that depends only on d', $\mathbf{W}'$ and k' (i.e. $d, \mathbf{W}$ and k). We will now set the parameter $t = t(d, k, \mathbf{W})$ to be simply $t_0 + 1$.

We also observe that our induction imposes that the polynomials

$$A_{\underline{\varepsilon},N,m}(m_1, ..., m_t) = \sum_{1\leq j\leq k'} p_{\underline{\varepsilon},j}(m_1, ..., m_t)u'_{j,m,N}$$

are non-constant and pairwise essentially distinct for any (non-zero) values of the leading vector $\{u'_{1,m,N}, ..., u'_{k',m,N}\}$ and that all the polynomials $p_{\underline{\varepsilon},j}$ are at most linear in each variable. In addition, we claim that

$$A_{\underline{\varepsilon},N,m}(m_1, ..., m_t) + A_{\underline{\varepsilon}',N,m}(m_1, ..., m_t) = A_{\underline{\varepsilon}+\underline{\varepsilon}',N}(m_1, ..., m_t) \text{ whenever } \underline{\varepsilon} + \underline{\varepsilon}' \in [[s]]. \tag{3.18}$$

(we have seen that all of the above are true in the linear case). These are the properties i)-iii) in Proposition 3.4.5.

All the $u'_{j,m,N}$ have the form described by Lemma 3.4.6. Therefore, we can write

$$p_{\underline{\varepsilon},j}(m_1, ..., m_t)u'_{j,m,N} = p'_{1,\underline{\varepsilon},\ell}(m, m_1, ..., m_t)u_{\ell,N} + p'_{2,\underline{\varepsilon},\ell'}(m_1, ..., m_t)u_{\ell',N}. \tag{3.19}$$

In order to describe the form of the new polynomials $p'_{1,\underline{\varepsilon},\ell}, p'_{2,\underline{\varepsilon},\ell'}$, we split into cases depending on the form of $u'_{j,m,N}$ (cf. Lemma 3.4.6):

a) If $u'_{j,m,N}$ is equal to some $u_{\ell,N}$ for $1 \leq \ell \leq k$, then we have $p'_{2,\underline{\varepsilon},\ell'} = 0$ and

$$p'_{1,\underline{\varepsilon},\ell}(m, m_1, ..., m_t) = p_{\underline{\varepsilon},j}(m_1, ..., m_t)$$

(thus $p'_{1,\underline{\varepsilon},\ell}(m_1, ..., m_t)$ is constant as a polynomial in m).

b) If $u'_{j,m,N}$ is equal to $dmu_{1,N}$ ($\ell = 1$), then we have again $p'_{2,\underline{\varepsilon},\ell'} = 0$ and

$$p'_{1,\underline{\varepsilon},1}(m, m_1, ..., m_t) = dmp_{\underline{\varepsilon},j}(m_1, ..., m_t).$$

c) In the final case that $u'_{j,m,N} = dmu_{1,N} + u_{\ell',N}$ for some $\ell' \neq 1$, then we have $p'_{2,\underline{\varepsilon},\ell'} = p_{\underline{\varepsilon},j}(m_1, ..., m_t)$ and

$$p'_{1,\underline{\varepsilon},1}(m, m_1, ..., m_t) = dmp_{\underline{\varepsilon},j}(m_1, ..., m_t).$$

Therefore, the new polynomials $p_{1,\underline{\varepsilon},\ell}$ and $p_{2,\underline{\varepsilon},\ell'}$ are at most linear in each of the variables $m_1, ..., m_t$, as well as the new variable m. By grouping the terms corresponding to the same $u_{\ell,N}$, we can rewrite

$$\sum_{1\leq r\leq k'} p_{\underline{\varepsilon},r}(m_1, ..., m_t)u'_{r,m,N} = \sum_{1\leq r\leq k} q_{\underline{\varepsilon},r}(m, m_1, ..., m_t)u_{r,N}$$

for some new polynomials $q_{\underline{\varepsilon},r}$.

Claim 1. *The new polynomials* $\sum_{1\leq r\leq k} q_{\underline{\varepsilon},r}(m, m_1, ..., m_t)u_{r,N}$ *satisfy conditions i), ii), iii) and iv) of Proposition 3.4.5, for any (non-zero) values of the* $u_{r,N}$.

Proof of the Claim. The fact that they are non-constant is trivial, since otherwise one of the polynomials

$$\sum_{1\leq r\leq k'} p_{\underline{\varepsilon},r}(m_1, ..., m_t)u'_{r,m,N}$$

would be constant, which is at odds with the induction hypothesis. Assume that condition ii) fails for two $\underline{\varepsilon_1}, \underline{\varepsilon_2} \in [[s]]$. Regarding these two polynomials as polynomials only in $(m_1, ..., m_t)$, (3.19) would give that the polynomials

$$\sum_{1\leq r\leq k'} p_{\underline{\varepsilon_1},r}(m_1, ..., m_t)u'_{r,m,N} \quad\text{and}\quad \sum_{1\leq r\leq k'} p_{\underline{\varepsilon_2},r}(m_1, ..., m_t)u'_{r,m,N},$$

are not essentially distinct, which is false by the induction hypothesis. Therefore, we have established both i) and ii).

Now, we want to prove an analogue of (3.18) for our new polynomials. But this follows by (3.19) (the new polynomials are just a rewritten form of the $A_{\underline{\varepsilon},N}$). This establishes that the new polynomials satisfy condition iii) in the statement of Proposition 3.4.5.

Finally, we are going to prove that the new polynomials $q_{\underline{\varepsilon},j}$ satisfy condition iv) of Proposition 3.4.5. Fix a $\underline{\varepsilon} \in [[s]]$. We will assume that all $q_{\underline{\varepsilon},j}$ are non-zero and we will show that they are linearly independent (if there are identically zero polynomials among the $q_{\underline{\varepsilon},j}$, we proceed similarly by ignoring these polynomials). It suffices to show that if $a_1, ..., a_k$ are real numbers, such that

$$a_1 q_{\underline{\varepsilon},1}(m, m_1, ..., m_t) + \cdots + a_k q_{\underline{\varepsilon},k}(m, m_1, ..., m_t)$$

is the zero polynomial, then all the numbers a_i are zero. Recalling the form of the $q_{\underline{\varepsilon},r}$, this becomes a linear combination of the form

$$a_1 P_{1,\underline{\varepsilon}}(m, m_1, ..., m_t) + \sum_{i\in I_1} b_i p_{\underline{\varepsilon},i}(m_1, ..., m_t) \tag{3.20}$$

for some $I_1 \subset \{1, 2, ..., k'\}$ and $b_i \in \{a_2, ..., a_k\}$[7]. In addition, the polynomial $P_{1,\underline{\varepsilon}}$ has the form

$$dm \sum_{i\in I_2} p_{\underline{\varepsilon},i} + \sum_{i\in I_3} p_{\underline{\varepsilon},i}$$

for some $I_2, I_3 \subset \{1, 2, ..., k'\}$ with $I_1 \cap I_2 = \emptyset$ and $I_1 \cap I_3 = \emptyset$. We argue by contradiction. For $m = 0$, the polynomial in (3.20) must be identically zero and this easily yields that all the b_i must be zero and that $a_1 \sum_{i\in I_3} p_{\underline{\varepsilon},i}$ is also the zero polynomial. The first relation implies that $a_2 = ... = a_k = 0$ by the induction hypothesis, while the second implies that either $a_1 = 0$ (in which case we are done), or $I_3 = \emptyset$ (since the $p_{\underline{\varepsilon},i}$ are linearly independent by the induction hypothesis). If $I_3 = \emptyset$, then (3.20) implies that the polynomial

$$a_1 dm \sum_{i\in I_2} p_{\underline{\varepsilon},i}$$

is the zero polynomial. This implies that $a_1 = 0$ or $I_2 = \emptyset$. However, we cannot have $I_2 = I_3 = \emptyset$, because that would imply that the polynomial $q_{\underline{\varepsilon},1}$ is identically zero, which is absurd (since we assumed that we have already discarded the zero polynomials among the $q_{\underline{\varepsilon},i}$). Our claim follows. □

Combining all of the above we rewrite (3.17) as

$$\sup_{|c_{n,N}|\leq 1} \Big\| \mathop{\mathbb{E}}_{0\leq n\leq L_N} c_{n,N} \Big(\prod_{i=1}^{k-1} T^{\lfloor p_{i,N}(n+m) - p_N(n) \rfloor} f_{i,N} \cdot T^{\lfloor p_{i,N}(n) - p_N(n) \rfloor} \overline{f_{i,N}} \Big) \cdot$$
$$T^{\lfloor p_{k,N}(t+m) - p_{k,N}(t) \rfloor} f_{k,N} \Big\|_{L^2(\mu)}^{2^{t_0+1}} \ll_{d,k,\mathbf{W}}$$
$$\frac{1}{M} + \sum_{\mathbf{h}\in[[Y_0]]} \mathop{\mathbb{E}}_{|m_1|,...,|m_t|\leq M} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} T^{\lfloor \sum_{1\leq r\leq k} q_{\underline{\varepsilon},r}(m,m_1,...,m_t) u_{r,N} \rfloor + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} f_{1,N}) d\mu \Big| + o_N(1).$$

We use the above bounds for all $-M \leq m \leq M$ in (3.16). The possible error coming from the bad values of the set Q can be absorbed by an $O_{d,k}(1/M)$ term. Finally, we get

$$\Big\| \mathop{\mathbb{E}}_{0\leq h\leq L_N} c_{n,N} \prod_{i=1}^{k} T^{\lfloor p_{i,N}(n) \rfloor} f_{i,N} \Big\|_{L^2(\mu)}^{2^t} \ll_{d,k,\mathbf{W}}$$
$$\frac{1}{M} + \sum_{\mathbf{h}\in[[Y_0]]} \mathop{\mathbb{E}}_{|m|,|m_1|,...,|m_t|\leq M} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} T^{\lfloor \sum_{1\leq r\leq k} q_{\underline{\varepsilon},r}(m,m_1,...,m_t) u_{r,N} \rfloor + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} f_{1,N}) d\mu \Big| + o_N(1),$$

which is what we wanted to show. □

[7] Observe that each one of the numbers $a_2, .., a_k$ appears in the set $\{b_i, i \in I_1\}$ (maybe with multiplicity), because we have assumed that each polynomial $q_{\underline{\varepsilon},\ell}$, $(\ell > 1)$ is not the trivial polynomial (otherwise, we ignore it).

3.5 Approximation of Hardy field functions

3.5.1 The sub-classes $S(a,k)$

We have already established Proposition 2.1.5 that can be used to approximate Hardy sequences in appropriate short intervals. In the proofs of the main theorems, we need to do the above approximation for several Hardy field functions in tandem. In order to achieve this, we will need to extend the aforementioned proposition to hold for several functions.

We introduce some terminology first. Let $a \in \mathcal{H}$ be a strongly non-polynomial Hardy function such that $a(t) \gg t^{\delta}$, for some $\delta > 0$. Namely, we exclude sub=fractional functions. For such a function a and $k \in \mathbb{N}$ sufficiently large (it is only required that $a^{(k)}(t) \to 0$), we define the subclass $S(a,k)$ of $\mathcal{H}$ as

$$S(a,k) = \{g \in \mathcal{H} : |a^{(k)}(t)|^{-\frac{1}{k}} \preceq g(t) \prec |a^{(k+1)}(t)|^{-\frac{1}{k+1}}\},$$

where we recall that the notation $g(t) \preceq f(t)$ means that the limit $\lim\limits_{t\to\infty} |f(t)/g(t)|$ is non-zero. Note that every $g \in S(a,k)$ is a sub-linear function, that is $g(t) \prec t$. Some very basic properties of the classes $S(a,k)$ are established in the following lemma.

Lemma 3.5.1. *Let $a \in \mathcal{H}$ be a strongly non-polynomial function with $a(t) \gg t^{\delta}$, for some $\delta > 0$.*
i) The class $S(a,k)$ is non-empty, for k sufficiently large.
ii) For any $0 < c < 1$ sufficiently close to 1, there exists $k_0 \in \mathbb{N}$, such that the function $t \to t^c$ of $\mathcal{H}$ belongs to $S(a,k_0)$.
iii) The class $S(a,k)$ does not contain all functions of the form $t \to t^c$, for c sufficiently close to 1.

Proof. i) This follows immediately from Proposition 2.1.5. We can actually show something stronger, namely, that if $a(t) \gg t^{\delta}$ for some $0 < \delta < 1$, then

$$\frac{|a^{(k+1)}(t)|^{-\frac{1}{k+1}}}{|a^{(k)}(t)|^{-\frac{1}{k}}} \gg t^{\frac{\delta}{k(k+1)}}, \tag{3.21}$$

which means that the functions at the "endpoints" of $S(a,k)$ differ by a fractional power. This last inequality follows by combining the relations

$$a^{(k)}(t) \gg t f^{(k+1)}(t) \qquad \text{and} \qquad a^{(k)}(t) \gg t^{\delta-k}.$$

ii) This is contained in Proposition 2.1.5.
iii) It suffices to show that there exists $c \in [0,1]$ such that $|a^{(k)}(t)|^{-\frac{1}{k}} \prec t^c$. For the sake of contradiction, we assume that this last inequality fails and use the lower bound from Lemma 2.1.3, to deduce that

$$t^{-kc} \gg a^{(k)}(t) \succ \frac{a(t)}{t^k \log^{2k} t}$$

for every $0 < c < 1$. This, implies that $a(t) \ll t^{k(1-c)} \log^{2k} t$ for all $0 < c < 1$, which contradicts the hypothesis that $a(t)$ is not sub-fractional. We remark in passing that this argument also indicates that the integer k can be made arbitrarily large by choosing c to be sufficiently close to 1. □

In essence, the claim implies that the classes $S(f,k)$ form a "partition" of the subclass

$$A = \{g(t) \gg t^c \colon\ \exists\ \delta > 0, \ \text{ with } g(t) \ll t^{1-\delta}\}$$

for some $c > 0$. That means that any sub-linear function that grows approximately as a (sufficiently large) fractional power must be contained in the union of the $S(a,k)$. This union however does not contain functions that are "logarithmically close" to linear functions, such as $t(\log t)^{-1}$. Although inaccurate, it is instructive to imagine the classes $S(a,k)$ as (disjoint) intervals on the real line. For example, if $S(a,k) = \{g(t)\colon \sqrt{t} \preceq g(t) \prec t^{2/3}\}$, then we can think that $S(a,k)$ is represented by the interval $[\frac{1}{2}, \frac{2}{3})$.

The following proposition relates the behavior of the subclasses $S(a,k)$ and $S(b,\ell)$ for different functions $a, b \in \mathcal{H}$.

Proposition 3.5.2. *For any two functions $a(t), b(t) \in \mathcal{H}$ as in Lemma 3.5.1 that also satisfy $b(t) \ll a(t)$, we have the following:*

i) The relation $S(a,k) = S(b,k)$ holds for some $k \in \mathbb{N}$ if and only if $a(t) \sim b(t)$.

ii) If $S(a,k) \cap S(b,\ell) \neq \emptyset$, then $k \geq \ell$. In addition, if the function $(a^{(k)}(t))^{-\frac{1}{k}}$ is contained in $S(b,\ell)$ and $a \not\sim b$, then $k \geq \ell+1$.

iii) There exist infinitely many pairs of integers (k,ℓ), such that $S(a,k) \cap S(b,\ell) \neq \emptyset$.

Proof. i) It is a straightforward application of L' Hospital's rule.

ii) Since the given intersection is non-empty, we must necessarily have $|b^{(\ell)}(t)|^{-\frac{1}{\ell}} \preceq |a^{(k+1)}(t)|^{-\frac{1}{k+1}}$. Suppose that $k < \ell$, so that we have the inequalities $|b^{(\ell)}(t)|^{-\frac{1}{\ell}} \preceq |a^{(k+1)}(t)|^{-\frac{1}{k+1}} \preceq |a^{(l)}(t)|^{-\frac{1}{\ell}}$, which implies that $a^{(l)}(t) \preceq b^{(l)}(t)$. Because we also have $b(t) \ll a(t)$, we can easily deduce that $a(t) \sim b(t)$ using the fact that both of these functions are strongly non-polynomial. Thus, the intersection $S(a,k) \cap S(b,\ell)$ is non-empty if and only if $k = \ell$, which is a contradiction.

For the proof of the second part, we use immediately the fact that $k \geq \ell$, which follows by the first part. Suppose that $k = l$ and we shall arrive at a contradiction. If $(a^{(k)}(t))^{-\frac{1}{k}} \in S(b,k)$, then, we must have $(a^{(k)}(t))^{-\frac{1}{k}} \succeq (b^{(k)}(t))^{-\frac{1}{k}}$, which implies that $b^{(k)}(t) \succeq a^{(k)}(t)$. This contradicts the assumption that $a(t) \succ b(t)$ (apply L' Hospital's rule k times).

iii) For any c close to 1, we can find k, such that the function t^c belongs to $S(a,k)$ (this follows from the second statement of Lemma 3.5.1) and similarly for the Hardy function b. Then, the intersection $S(a,k) \cap S(b,\ell)$ is non-empty. Taking $c \to 1$ from below and using the third statement of Lemma 3.5.1, we can find infinitely many such pairs. □

Remark. It is straightforward to generalize the third statement of the above proposition to the case of k distinct functions $a_1, ..., a_k$ in $\mathcal{H}$. We will use this observation in our arguments to find a function L in the intersection of these classes. Note that our previous discussion implies that for such a function $L(t)$, all the involved functions $a_1, ..., a_k$ will have a polynomial expansion on intervals of the form $[N, N + L(N)]$ and this will play a crucial role in our approximations.

3.5.2 The subclasses $S_{sml}(a,k)$

We can similarly define analogs of the classes $S(a,k)$ for functions with small growth rate, that is sub-fractional functions. Let $a \in \mathcal{H}$ be a sub-fractional function such that $\log t \prec a(t)$. If $k \geq 1$, we can define the class

$$S_{sml}(f,k) = \{g \in \mathcal{H}\colon\ |f^{(k)}(t)|^{-\frac{1}{k}} \preceq g(t) \prec |f^{(k+1)}(t)|^{-\frac{1}{k+1}}\}.$$

The properties of Proposition 3.5.2 proven for the classes $S(a,k)$ are carried verbatim to this new setting. The major difference is that now every function $g \in S_{sml}(f,k)$ dominates all functions of the form $t^{1-\delta}$ for $\delta > 0$ (an example is the function $t/\log t$). In particular, $S_{sml}(f,k)$ has trivial intersection with the classes $S(h,\ell)$ defined above for any integers k,ℓ and appropriate functions f,h.

As an example, let us consider a fractional power t^δ with $0 < \delta < 1$ and two functions $f, g \in \mathcal{H}$ such that $f(t) \gg t^\varepsilon$ for some $\varepsilon > 0$, while $\log t \prec g(t)$ and g is sub-fractional. A typical case is the pair $(t^{3/2}, \log^2 t)$. We know that if δ is close enough to 1, then the function t^δ will belong to $S(f,k)$ for some $k \in \mathbb{N}$. Using approximations similar to the ones in the previous subsection, we can see that the sequence $f(n)$ becomes a polynomial sequence of degree k on intervals of the form $[N, N+N^\delta]$. On the other hand, the sequence $g(n)$, restricted to the same interval, is $o_N(1)$ close to the value $g(N)$, which means that it is "essentially" constant on this interval. This difference in behavior leads to some added complexity in our proofs, since some of our functions may be approximated by polynomials, while other functions become constant.

On the other hand, a function $a \in \mathcal{H}$ with $a(t) \ll \log t$, when restricted to intervals of the form $[N, N + L(N)]$, is $o_N(1)$-close to the value $f(N)$ for any sub-linear function $L(t)$. Functions of this form always collapse to a constant when restricted to intervals of the above form.

We will not use of the classes $\mathcal{S}_{sml}(a,k)$ in the main proof. These classes need to be used only in the case when all the functions in Proposition 3.1.1 are sub-fractional. The arguments are very similar (and simpler) to the case when we have both sub-fractional and super-fractional functions, with the only difference being the initial choice of the short interval. We only mention them in order to highlight that they share the same properties as the $\mathcal{S}(a,k)$.

3.6 The sub-linear plus polynomial case

In this section, we establish a particular case of Proposition 3.1.1, which we shall also use in the general case in the next section. Let $\mathcal{S}$ denote the subset of $\mathcal{H}$ that contains the functions with sub-linear growth rate and $\mathcal{P}\subseteq\mathcal{H}$ denotes the collection of polynomials with real coefficients. Then, we let $\mathcal{S}+\mathcal{P}$ denote the collection of functions that can be written as a sum of a function in $\mathcal{S}$ and a function in $\mathcal{P}$ (or equivalently, linear combinations of functions in $\mathcal{S}$ and $\mathcal{P}$).

Let $a_1,...,a_k$ be a collection of functions in $\mathcal{S}+\mathcal{P}$. Then, we can write $a_i=u_i+p_i$, where $u_i\in\mathcal{S}$ and p_i is a polynomial. We will also define the *degree* and *type* of the collection $a_1,a_2,...,a_k$ using a similar notion to the degree and type of a polynomial family defined in the previous section. More precisely, since we do not impose that the polynomials $p_1,...,p_k$ are essentially distinct, we choose a maximal subset of the polynomials p_i consisting of non-constant and essentially distinct polynomials and we define the degree and type of the collection $a_1,...,a_k$ to be the degree and type of this new subfamily of polynomials, respectively. Similarly, we define the leading vector of $a_1,...,a_k$ as the leading vector of the maximal subfamily that we defined above. We can always choose this maximal subset to contain the polynomial p_1. We define the cardinality of this new maximal subset to be the *size* of the collection $a_1,...,a_k$.

Proposition 3.6.1. *Let M be a positive integer and let $a_1,...,a_k$ be a collection of functions in $\mathcal{S}+\mathcal{P}$ with degree d, type $\mathbf{W}$ and size $k'\le k$. Let $(c_1,...,c_{k'})$ be the leading vector of the family $\{a_1,...,a_k\}$. In addition, assume that $a_1(t)\succ\log t$ and $a_1(t)-a_j(t)\succ\log t$ for $j\neq 1$. Then, there exist positive integer s,t, a finite set Y of integers and real polynomials $p_{\underline{\varepsilon},j}$ in t variables, where $\underline{\varepsilon}\in[[s]]$ and $1\le j\le k$, all depending only on $d,k',\mathbf{W}$, such that, for any measure preserving system (X,μ,T) and function $f_1\in L^\infty(\mu)$ bounded by 1, we have*

$$\sup_{\|f_2\|_\infty,...,\|f_k\|_\infty\le 1}\ \sup_{|c_n|\le 1}\Big\|\underset{1\le n\le N}{\mathbb{E}} c_n\, T^{\lfloor a_1(n)\rfloor}f_1\cdot...\cdot T^{\lfloor a_k(n)\rfloor}f_k\Big\|^{2^t}_{L^2(\mu)}\ll_{d,k,k',\mathbf{W}}$$
$$\frac{1}{M}+\sum_{\mathbf{h}\in Y^{[[s]]}}\underset{m\in[-M,M]^t}{\mathbb{E}}\Big|\!\Big|\!\Big|\prod_{\underline{\varepsilon}\in[[s]]}T^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor+h_{\underline{\varepsilon}}}f_1\Big|\!\Big|\!\Big|_{2k+1}+o_N(1)\quad(3.22)$$

where

$$A_{\underline{\varepsilon}}(\mathbf{m})=\sum_{j=1}^{k'}p_{\underline{\varepsilon},j}(\mathbf{m})c_j.$$

are pairwise essentially distinct polynomials.

Observe that the iterates inside the seminorm in (3.22) are real polynomials in several variables. We can take $M\to+\infty$ and expand these seminorms to arrive at an iterated limit of polynomial averages. It is possible to bound these averages by a suitable seminorm of the function f_1 using the results in [38] and get a simpler bound in (3.22). This necessitates that we substitute the $O_{d,k,k',\mathbf{W}}(1)$ implicit constant by an $O_{a_1,...,a_k}(1)$ constant and this is insufficient for our purposes in the next section, where we will have to apply Proposition 3.6.1 for several collections of functions simultaneously. However, in view of the above discussion, we can deduce the following:

Corollary 3.6.2. *Let $a_1,...,a_k$ be a collection of functions in $\mathcal{S}+\mathcal{P}$ such that $a_1(t)\succ\log t$ and $a_1(t)-a_j(t)\succ\log t$ for $j\neq 1$. Then, there exists a positive integer s such that, for any measure preserving system (X,μ,T) and 1-bounded function $f_1\perp Z_s(X)$, we have*

$$\lim_{N\to+\infty}\ \sup_{\|f_2\|_\infty,...,\|f_k\|_\infty\le 1}\ \sup_{|c_n|\le 1}\Big\|\underset{1\le n\le N}{\mathbb{E}} c_n\, T^{\lfloor a_1(n)\rfloor}f_1\cdot...\cdot T^{\lfloor a_k(n)\rfloor}f_k\Big\|_{L^2(\mu)}=0.$$

We analyze the conditions imposed on the functions $a_1, ..., a_k$ more closely: write each function a_i in the form $a_i(t) = u_i(t) + p_i(t)$, where $u_i \in \mathcal{S}$ and $p_i \in \mathcal{P}$. The condition $a_1(t) \succ \log t$ implies that either $u_1(t) \succ \log t$ or $p_1(t)$ is a non-constant polynomial. Similarly, the second condition implies that either $u_1(t) - u_i(t) \succ \log t$ or $p_1(t) - p_j(t)$ is a non-constant polynomial.

Furthermore, we can make one more reduction. Writing again $a_i(t) = u_i(t) + p_i(t)$ as above and using the same argument as in the discussion following the statement of Proposition 3.1.1), we may assume that the function u_1 has the largest growth rate among the functions u_i.

In order to establish the main result of this section, we will also use the following proposition, which is special case of Proposition 3.6.1.

Proposition 3.6.3. *Let $a_1, ..., a_k$ be sub-linear functions in $\mathcal{H}$ and assume that all the functions $a_1, a_1 - a_2, ..., a_1 - a_k$ dominate $\log t$. Then, for any measure preserving system (X, μ, T) and function $f_1 \in L^\infty(\mu)$ bounded by 1, we have*

$$\limsup_{N\to+\infty} \sup_{\|f_2\|_\infty,...,\|f_k\|_\infty \leq 1} \sup_{|c_n|\leq 1} \Big\| \mathop{\mathbb{E}}_{1\leq n\leq N} c_n\, T^{\lfloor a_1(n)\rfloor} f_1 \cdot ... \cdot T^{\lfloor a_k(n)\rfloor} f_k \Big\|_{L^2(\mu)} \ll_k \, \|f_1\|_{2k}. \tag{3.23}$$

Remark. The proof that Proposition 3.6.3 implies Proposition 3.6.1 corresponds to Step 1 in example b) of this chapter, while the proof of Proposition 3.6.3 corresponds to step 2 of the same example.

Proof that Proposition 3.6.3 implies Proposition 3.6.1. First of all, we write each $a_i(t)$ in the form $u_i(t) + p_i(t)$ as we discussed above. Our main tool will be to use Lemma 3.2.2 in order to reduce our problem to studying averages on small intervals, where the sublinear functions u_i will have a constant integer part.

Suppose that not all of the polynomials $p_1(t), ..., p_k(t)$ are constant, since that case follows from Proposition 3.6.3 (that means the family has degree ≥ 1). We can assume, without loss of generality, that $p_i(0) = 0$ for all i (the constant terms can be absorbed by the functions u_i). Therefore, let $L(t) \in \mathcal{H}$ be a sub-linear function to be chosen later. In addition, we choose functions $f_{2,N}, ..., f_{k,N}$ so that the average in the left-hand side of (3.22) is $1/N$ close to the supremum. We want to bound

$$\mathop{\mathbb{E}}_{1\leq r\leq R} \sup_{|c_{n,r}|\leq 1} \Big\| \mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} c_{n,r}\, T^{\lfloor u_1(n)+p_1(n)\rfloor} f_1 \cdot ... \cdot T^{\lfloor u_k(n)+p_k(n)\rfloor} f_{k,R} \Big\|^{2^t}_{L^2(\mu)}$$

for some integer parameter t, which we will choose later to depend only on the quantities $d, k', \mathbf{W}$ (thus, when applying Lemma 3.2.1 below to remove the error terms in the iterates, we will always have that the implicit constant depends only on $d, k', \mathbf{W}$).

Recall that we have reduced our problem to the case that the function u_1 has the largest growth rate among the functions u_i. Now, we want to choose the sub-linear function $L(t) \in \mathcal{H}$ so that the functions $u_i(n)$ restricted to the interval $[r, r + L(r)]$ become very close to the value $u_i(r)$. To achieve this, it suffices to take $L(t) \in \mathcal{H}$ such that

$$1 \prec L(t) \prec (u_1'(t))^{-1}.$$

To see that such a function exists, we only need to show that $(u_1'(t))^{-1} \succ 1$ which follows easily from the fact that $u_1(t) \prec t$. Observe that for every $i \in \{1, 2, ..., k\}$ we must have $L(t) \prec (u_i'(t))^{-1}$, since u_1 has maximal growth among the functions u_i. For every $n \in [r, r + L(r)]$, we observe that

$$|u_i(n) - u_i(r)| \leq (n - r) \max_{x\in[r,r+L(r)]} |u_i'(x)|.$$

Since $|u_i'(t)| \searrow 0$, we have that for r large enough

$$|u_i(n) - u_i(r)| \leq L(r)u_i'(r) = o_r(1), \qquad n \in [r, r + L(r)].$$

Therefore, for r sufficiently large we have

$$\lfloor u_i(n) + p_i(n)\rfloor = \lfloor u_i(r)\rfloor + \lfloor p_i(n)\rfloor + e_{i,n}, \qquad n \in [r, r + L(r)],$$

where $e_{i,n} \in \{0, \pm 1, \pm 2\}$. Therefore, our original problem reduces to bounding the quantity

$$\mathop{\mathbb{E}}_{1\leq r\leq R} \sup_{|c_{n,r}|\leq 1} \Big\| \mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} c_{n,r}\, T^{\lfloor u_1(r)\rfloor+\lfloor p_1(n)\rfloor+e_{1,n}} f_1 \cdot \ldots \cdot T^{\lfloor u_k(r)\rfloor+\lfloor p_k(n)\rfloor+e_{k,n}} f_{k,R} \Big\|^{2^t}_{L^2(\mu)}. \tag{3.24}$$

Using Lemma 3.2.1, we may reduce to the case that the error terms $e_{i,n}$ in the iterates are all equal to zero.

Let S be the set of those $i \in \{1, ..., k\}$ for which the polynomial $p_i(t)$ is equal to the polynomial $p_1(t)$. Reordering, if necessary, we may assume that $S = \{1, ..., k_0\}$ for some $k_0 \leq k$. Note that the original condition then implies that $u_1(t) - u_i(t) \succ \log t$ for each $2 \leq i \leq k_0$. We rewrite (3.24) as

$$\begin{aligned} &\mathop{\mathbb{E}}_{1\leq r\leq R} \sup_{|c_{n,r}|\leq 1} \Big\| \mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} c_{n,r}\, T^{\lfloor p_1(n)\rfloor} (\prod_{i=1}^{k_0} T^{\lfloor u_i(r)\rfloor} f_{i,R}) \prod_{i=k_0+1}^{k} T^{\lfloor u_i(r)\rfloor+\lfloor p_i(n)\rfloor} f_{i,R} \Big\|^{2^t}_{L^2(\mu)} = \\ &\mathop{\mathbb{E}}_{1\leq r\leq R} \sup_{|c_{h,r}|\leq 1} \Big\| \mathop{\mathbb{E}}_{0\leq h\leq L(r)} c_{h,r}\, T^{\lfloor p_1(r+h)\rfloor} (\prod_{i=1}^{k_0} T^{\lfloor u_i(r)\rfloor} f_{i,R}) \prod_{i=k_0+1}^{k} T^{\lfloor u_i(r)\rfloor+\lfloor p_i(r+h)\rfloor} f_{i,R} \Big\|^{2^t}_{L^2(\mu)} \leq \\ &\mathop{\mathbb{E}}_{1\leq r\leq R} \sup_{\|f_{k_0+1}\|_\infty,\ldots,\|f_k\|_\infty\leq 1} \sup_{|c_{h,r}|\leq 1} \Big\| \mathop{\mathbb{E}}_{0\leq h\leq L(r)} c_{h,r}\, T^{\lfloor p_1(r+h)\rfloor} (\prod_{i=1}^{k_0} T^{\lfloor u_i(r)\rfloor} f_{i,R}) \prod_{i=k_0+1}^{k} T^{\lfloor p_i(r+h)\rfloor} f_i \Big\|^{2^t}_{L^2(\mu)}, \end{aligned} \tag{3.25}$$

where $f_{1,R} = f_1$. We also write $F_{r,R} := \prod_{i=1}^{k_0} T^{\lfloor u_i(r)\rfloor} f_{i,R}$ for brevity.

We can assume that the polynomials $p_i(r+h)$ are non-constant (otherwise, we just ignore the corresponding iterate in the last average). In addition, we may assume that they are pairwise essentially distinct, because if two polynomials are equal, we can combine both of these iterates into a single iterate (this operation does not change the type or leading vector of the given collection of functions). Note that under these assumptions the family of polynomials

$$P_r = \{p_1(r+h), p_{k_0+1}(r+h), ..., p_k(r+h)\}$$

is a nice family of polynomials[8] in the variable h (the leading coefficients of the polynomials and their pairwise differences are all constant sequences) and has type and leading vector equal to that of the original collection $\{p_1, ..., p_k\}$. Therefore, we can apply Proposition 3.4.5: there exist positive integers t_0 and s, a finite set Y of integers and polynomials $p_{\underline{\varepsilon},j}$ where $\underline{\varepsilon} \in [[s]]$ and $1 \leq j \leq k$ such that

$$\begin{aligned} \sup_{\|f_{k_0+1}\|_\infty,\ldots,\|f_k\|_\infty\leq 1} \sup_{|c_{h,r}|\leq 1} \Big\| \mathop{\mathbb{E}}_{0\leq h\leq L(r)} c_{h,r}\, T^{\lfloor p_1(r+h)\rfloor} F_{r,R} \prod_{i=k_0+1}^{k} T^{\lfloor p_i(r+h)\rfloor} f_i \Big\|^{2^{t_0}}_{L^2(\mu)} \ll_{d,k',\mathbf{W}} \\ \frac{1}{M} + \sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^{t_0}} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} T^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} F_{r,R})\, d\mu \Big| + o_r(1), \end{aligned} \tag{3.26}$$

where

$$A_{\underline{\varepsilon}}(\mathbf{m}) = \sum_{1\leq j\leq k'} p_{\underline{\varepsilon},j}(\mathbf{m}) c_j$$

and $(c_1, ..., c_{k'})$ is the leading vector of the initial family (here we have $k' \leq k - k_0 + 1$).

Using this in (3.25) with $t = t_0$ (which depends only on $d, k', \mathbf{W}$ as we claimed in the beginning), we deduce that our original average is bounded by $O_{d,k,k',\mathbf{W}}(1)$ times

$$\frac{1}{M} + \mathop{\mathbb{E}}_{1\leq r\leq R} \sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^{t_0}} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} T^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} F_{r,R})\, d\mu \Big| + o_R(1).$$

[8] There is the possibility that the polynomial $p_1(n)$ is constant (and so is the polynomial $p_1(r+h)$) or that it does not have maximal degree (which would prevent the use of Proposition 3.4.5, which was stated for ordered polynomial families). However, since we have assumed that not all of the polynomials p_i are constant, then we can use the same argument in the discussion after Proposition 3.1.1 (where we reduced our problem to the case that the first function has maximal growth rate) to replace the polynomial $p_1(r+h)$ by $p_1(r+h) - p_i(r+h)$ for a non-constant polynomial $p_i(r+h)$ among $p_{k_0+1}(r+h), ..., p_k(r+h)$.

Using the definition of $F_{r,R}$, we rewrite this as

$$\frac{1}{M}+\underset{1\le r\le R}{\mathbb{E}}\sum_{\mathbf{h}\in Y^{[[s]]}}\underset{\mathbf{m}\in[-M,M]^{t_0}}{\mathbb{E}}\Big|\int T^{\lfloor u_1(r)\rfloor}\big(\prod_{\underline{\varepsilon}\in[[s]]}T^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor+h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}f_1)\big)$$
$$\prod_{i=2}^{k_0}T^{\lfloor u_i(r)\rfloor}\big(\prod_{\underline{\varepsilon}\in[[s]]}T^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor+h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}f_{i,R})\big)\,d\mu\Big|+o_R(1).$$

Now, we consider two cases:

<u>Case 1</u>: Firstly, assume that $k_0=1$. Then, the above quantity can be rewritten as

$$\frac{1}{M}+\underset{1\le r\le R}{\mathbb{E}}\sum_{\mathbf{h}\in Y^{[[s]]}}\underset{\mathbf{m}\in[-M,M]^{t_0}}{\mathbb{E}}\Big|\int T^{\lfloor u_1(r)\rfloor}\big(\prod_{\underline{\varepsilon}\in[[s]]}T^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor+h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}f_1)\big)\,d\mu\Big|+o_R(1)=$$
$$\frac{1}{M}+\sum_{\mathbf{h}\in Y^{[[s]]}}\underset{\mathbf{m}\in[-M,M]^{t_0}}{\mathbb{E}}\Big|\int\prod_{\underline{\varepsilon}\in[[s]]}T^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor+h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}f_1)\big)\,d\mu\Big|+o_R(1).$$

The result follows immediately, since

$$\Big|\int\prod_{\underline{\varepsilon}\in[[s]]}T^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor+h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}f_1)\big)\,d\mu\Big|\le\Big\|\prod_{\underline{\varepsilon}\in[[s]]}T^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor+h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}f_1)\big)\Big\|_{2k+1}.$$

<u>Case 2</u>: Assume that $k_0>1$ and we want to bound

$$\frac{1}{M}+\underset{1\le r\le R}{\mathbb{E}}\sum_{\mathbf{h}\in Y^{[[s]]}}\underset{\mathbf{m}\in[-M,M]^{t_0}}{\mathbb{E}}\Big|\int T^{\lfloor u_1(r)\rfloor}\big(\prod_{\underline{\varepsilon}\in[[s]]}T^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor+h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}f_1)\big)$$
$$\prod_{i=2}^{k_0}T^{\lfloor u_i(r)\rfloor}\big(\prod_{\underline{\varepsilon}\in[[s]]}T^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor+h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}f_{i,R})\big)\,d\mu\Big|+o_R(1).\quad(3.27)$$

Our original hypothesis implies that the functions u_1-u_i (where $2\le i\le k_0$) dominate $\log t$. Since u_i was assumed in the beginning to have the biggest growth rate among the functions u_i, we must also have $u_1(t)\succ\log t$.

We take the limit as $R\to+\infty$ and rewrite the quantity in (3.27) as

$$\frac{1}{M}+\sum_{\mathbf{h}\in Y^{[[s]]}}\underset{\mathbf{m}\in[-M,M]^{t}}{\mathbb{E}}\limsup_{R\to+\infty}\underset{1\le r\le R}{\mathbb{E}}\Big|\int T^{\lfloor u_1(r)\rfloor}\big(\prod_{\underline{\varepsilon}\in[[s]]}T^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor+h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}f_1)\big)$$
$$\prod_{i=2}^{k_0}T^{\lfloor u_i(r)\rfloor}\big(\prod_{\underline{\varepsilon}\in[[s]]}T^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor+h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}f_{i,R})\big)\,d\mu\Big|.$$

Applying the Cauchy-Schwarz inequality, we deduce that

$$\underset{1\le r\le R}{\mathbb{E}}\Big|\int T^{\lfloor u_1(r)\rfloor}\big(\prod_{\underline{\varepsilon}\in[[s]]}T^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor+h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}f_1)\big)\prod_{i=2}^{k_0}T^{\lfloor u_i(r)\rfloor}\big(\prod_{\underline{\varepsilon}\in[[s]]}T^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor+h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}f_{i,R})\big)\,d\mu\Big|\le$$
$$\Big(\underset{1\le r\le R}{\mathbb{E}}\int S^{\lfloor u_1(r)\rfloor}\big(\prod_{\underline{\varepsilon}\in[[s]]}S^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor+h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}F_1)\big)$$
$$\prod_{i=2}^{k_0}S^{\lfloor u_i(r)\rfloor}\big(\prod_{\underline{\varepsilon}\in[[s]]}S^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor+h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}F_{i,R})\big)\,d(\mu\times\mu)\Big)^{1/2},$$

where $S = T\times T$, $F_1 = \overline{f_1}\otimes f_1$ and $F_{i,R} = \overline{f_{i,R}}\otimes f_{i,R}$. A final application of the Cauchy-Schwarz inequality bounds the last quantity by

$$\Big\|\mathop{\mathbb{E}}_{1\leq r\leq R} S^{\lfloor u_1(r)\rfloor}\Big(\prod_{\underline{\varepsilon}\in[[s]]} S^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}F_1)\Big)\prod_{i=2}^{k_0} S^{\lfloor u_i(r)\rfloor}\Big(\prod_{\underline{\varepsilon}\in[[s]]} S^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}F_{i,R})\Big)\Big\|_{L^2(\mu\times\mu)}^{1/2}.$$

Applying Proposition 3.6.3, we deduce that the lim sup of this last average is bounded by $O_{k_0}(1)$ (which is $O_k(1)$) times

$$\Big|\!\Big|\!\Big|\prod_{\underline{\varepsilon}\in[[s]]} S^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}F_1)\Big|\!\Big|\!\Big|_{2k_0,T\times T} \leq \Big|\!\Big|\!\Big|\prod_{\underline{\varepsilon}\in[[s]]} S^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}F_1)\Big|\!\Big|\!\Big|_{2k,T\times T}.$$

Our original problem reduces to bounding

$$\frac{1}{M} + \sum_{\mathbf{h}\in Y^{[[s]]}}\mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t}\Big|\!\Big|\!\Big|\prod_{\underline{\varepsilon}\in[[s]]} S^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}F_1)\Big|\!\Big|\!\Big|_{2k,T\times T}^{1/2} + o_R(1),$$

which is smaller than

$$\frac{1}{M} + \sum_{\mathbf{h}\in Y^{[[s]]}}\mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t}\Big|\!\Big|\!\Big|\prod_{\underline{\varepsilon}\in[[s]]} T^{\lfloor A_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}f_1)\Big|\!\Big|\!\Big|_{2k+1,T} + o_R(1)$$

and the conclusion follows. □

Proof of Proposition 3.6.3. Using the arguments after the statement of Proposition 3.1.1, we may reduce to the case that $a_1(t)$ has maximal growth rate among $a_1, ..., a_k$.

We induct on k. In the base case of the induction, we want to show that

$$\limsup_{N\to+\infty}\sup_{|c_n|\leq 1}\Big\|\mathop{\mathbb{E}}_{1\leq n\leq N} c_n T^{\lfloor a_1(n)\rfloor} f_1\Big\|_{L^2(\mu)} \ll |\!|\!| f_1 |\!|\!|_2.$$

Due to Lemma 3.2.2, it suffices to show that

$$\limsup_{N\to+\infty}\sup_{|c_{n,N}|\leq 1}\Big\|\mathop{\mathbb{E}}_{N\leq n\leq N+L(N)} c_{n,N} T^{\lfloor a_1(n)\rfloor} f_1\Big\|_{L^2(\mu)} \ll |\!|\!| f_1 |\!|\!|_2 \tag{3.28}$$

for some suitable sub-linear function $L(t)\in\mathcal{H}$. Since $a_1(t)\succ \log t$, we conclude that

$$|a_1'(t)|^{-1} \prec |a_1''(t)|^{-1/2}$$

by a simple application of L' Hospital's rule. We choose the function $L(t)$ to satisfy

$$|a_1'(t)|^{-1} \prec L(t) \prec |a_1''(t)|^{-1/2}.$$

Therefore, for every $n\in[N, N+L(N)]$, we can write

$$a_1(n) = a_1(N) + (n-N)a_1'(N) + o_N(1),$$

which in turn implies that, for N sufficiently large, we can write

$$\lfloor a_1(n)\rfloor = \lfloor a_1(N) + (n-N)a_1'(N)\rfloor + e_{n,N},$$

where $e_{n,N}\in\{0,\pm 1\}$. Substituting this in (3.28), we want to prove that

$$\limsup_{N\to+\infty}\sup_{|c_{n,N}|\leq 1}\Big\|\mathop{\mathbb{E}}_{N\leq n\leq N+L(N)} c_{n,N} T^{\lfloor a_1(N)+(n-N)a_1'(N)\rfloor + e_{n,N}} f_1\Big\|_{L^2(\mu)} \ll |\!|\!| f_1 |\!|\!|_2.$$

Using Lemma 3.2.1, we can reduce our problem to

$$\limsup_{N\to+\infty}\sup_{|c_{h,N}|\leq 1}\Big\|\mathop{\mathbb{E}}_{0\leq h\leq L(N)} c_{h,N} T^{\lfloor a_1(N)+ha_1'(N)\rfloor} f_1\Big\|_{L^2(\mu)} \ll |\!|\!| f_1 |\!|\!|_2.$$

This bound can be proven using the change of variables trick that we have seen in the first example in Section 3.1. However, we will establish our assertion with a slightly quicker argument below.

We shall apply the van der Corput inequality. We fix a positive integer M and choose the quantity $M_N=\lfloor |M/a_1'(N)|\rfloor$. It is easy to check that $M_N\prec L(N)$, since $L(N)|a_1'(N)|\to+\infty$. Therefore, we can apply the van der Corput inequality to deduce that

$$\Big\|\mathop{\mathbb{E}}_{0\le h\le L(N)} c_{h,N}T^{\lfloor a_1(N)+ha_1'(N)\rfloor}f_1\Big\|_{L^2(\mu)}^2\ll$$
$$\frac{1}{M_N}+\mathop{\mathbb{E}}_{|m|\le M_N}\Big|\mathop{\mathbb{E}}_{0\le h\le L(N)}\overline{c_{h,N}}c_{h+m,N}\int T^{\lfloor a_1(N)+ha_1'(N)\rfloor}\overline{f_1}\cdot T^{\lfloor a_1(N)+(h+m)a_1'(N)\rfloor}f_1\,d\mu\Big|+o_N(1),$$

where the implied constant is absolute (and does not depend on M). We write

$$\lfloor a_1(N)+(h+m)a_1'(N)\rfloor=\lfloor a_1(N)+ha_1'(N)\rfloor+\lfloor ma_1'(N)\rfloor+e_{m,h,N},$$

where $e_{m,h,N}\in\{0,\pm1\}$. We rewrite the double average in the middle as

$$\mathop{\mathbb{E}}_{|m|\le M_N}\Big|\mathop{\mathbb{E}}_{0\le h\le L(N)}\overline{c_{h,N}}c_{h+m,N}\int\overline{f_1}\cdot T^{\lfloor ma_1'(N)\rfloor+e_{m,h,N}}f_1\,d\mu\Big|\le$$
$$\sum_{z\in\{0,\pm1\}}\mathop{\mathbb{E}}_{|m|\le M_N}\Big|\int\overline{f_1}\cdot T^{\lfloor ma_1'(N)\rfloor+z}f_1\,d\mu\Big|.$$

However, note that $|ma_1'(N)|\le M_N|a_1'(N)|\le M$. Thus, for any $z\in\{0,\pm1\}$, we have

$$\mathop{\mathbb{E}}_{|m|\le M_N}\Big|\int\overline{f_1}\cdot T^{\lfloor ma_1'(N)\rfloor+z}f_1\,d\mu\Big|=\frac{2M+1}{2M_N+1}\mathop{\mathbb{E}}_{|m'|\le M}p_N(m')\Big|\int\overline{f_1}\cdot T^{m'+z}f_1\,d\mu\Big|,$$

where $p_N(m')=\#\{m\in\mathbb{N}\colon \lfloor ma_1'(N)\rfloor=m'\}$. Since $a_1'(N)\to0$, we can easily see that for N large enough, we must have

$$p_N(m')\le\Big|\frac{1}{a_1'(N)}\Big|.$$

Therefore, we have

$$\frac{2M+1}{2M_N+1}\mathop{\mathbb{E}}_{|m'|\le M}p_N(m')\Big|\int\overline{f_1}\cdot T^{m'+z}f_1\,d\mu\Big|\le\frac{(2M+1)}{(2M_N+1)|a_1'(N)|}\mathop{\mathbb{E}}_{|m'|\le M}\Big|\int\overline{f_1}\cdot T^{m'+z}f_1\,d\mu\Big|\ll$$
$$\mathop{\mathbb{E}}_{|m'|\le M}\Big|\int\overline{f_1}\cdot T^{m'+z}f_1\,d\mu\Big|.$$

Thus, the square of our original average is $O(1)$ times

$$\sum_{z\in\{0,\pm1\}}\mathop{\mathbb{E}}_{|m'|\le M}\Big|\int\overline{f_1}\cdot T^{m'+z}f_1\,d\mu\Big|+o_N(1)$$

for some implied constant that does not depend on the original integer M. Therefore, we take first $N\to+\infty$ and then $M\to+\infty$ and use the Cauchy-Schwarz inequality to reach the conclusion. This establishes the base case of the induction.

Now assume the claim has been established for all positive integers less than or equal to $k-1$ (for some $k\ge2$). We prove that it holds for k as well. Since we have assumed that a_1 has maximal growth rate, we may reorder the given functions so that we have $a_1(t)\gg\cdots\gg a_k(t)$. Let $k_0\le k$ be the largest integer, such that the function a_{k_0} has the same growth rate as $a_1(t)$. This means that all the functions $a_1,\dots,a_{k_0}$ have the same growth rate. We rewrite our average in (3.23) as

$$\sup_{\|f_2\|_\infty,\dots,\|f_k\|_\infty\le1}\ \sup_{|c_n|\le1}\Big\|\mathop{\mathbb{E}}_{1\le n\le N}c_n\prod_{i=1}^{k_0}T^{\lfloor(a_i(n)-a_{k_0}(n))+a_{k_0}(n)\rfloor}f_i\cdot\prod_{i=k_0+1}^{k}T^{\lfloor a_i(n)\rfloor}f_i\Big\|_{L^2(\mu)}=$$
$$\sup_{\|f_2\|_\infty,\dots,\|f_k\|_\infty\le1}\ \sup_{|c_n|\le1}\Big\|\mathop{\mathbb{E}}_{1\le n\le N}c_n\,T^{\lfloor a_{k_0}(n)\rfloor}\Big(\prod_{i=1}^{k_0}T^{\lfloor(a_i(n)-a_{k_0}(n))\rfloor+e_{i,n}}f_i\Big)\prod_{i=k_0+1}^{k}T^{\lfloor a_i(n)\rfloor}f_i\Big\|_{L^2(\mu)}\quad(3.29)$$

for some $e_{i,n} \in \{0, \pm 1\}$. Using Lemma 3.2.1, we may reduce our problem to the case that all the error terms $e_{i,n}$ are zero. Note that the function $a_{k_0}(n)$ dominates each one of the functions $a_1 - a_{k_0}, ..., a_{k_0-1} - a_{k_0}$, as well as the functions $a_i, i \geq k_0$. Now, we choose sequences of functions $f_{2,N}, ..., f_{k,N}$ so that the above average is $1/N$ close to the supremum (we also write $f_{1,N} = f_1$). In addition, we use Lemma 3.2.2and, thus, it is sufficient to show that

$$\limsup_{R\to+\infty} \underset{1\leq r\leq R}{\mathbb{E}} \sup_{|c_{n,r}|\leq 1} \Big\| \underset{r\leq n\leq r+L(r)}{\mathbb{E}} c_{n,r}\, T^{\lfloor a_{k_0}(n)\rfloor} \Big(\prod_{i=1}^{k_0} T^{\lfloor (a_i(n)-a_{k_0}(n))\rfloor} f_{i,R}\Big) \prod_{i=k_0+1}^{k} T^{\lfloor a_i(n)\rfloor} f_{i,R}\Big\|_{L^2(\mu)} \ll_k \|\!|f_1|\!\|_{2k} \quad (3.30)$$

for a sub-linear function $L(t) \in \mathcal{H}$ that we shall choose momentarily. Namely, we choose the function $L \in \mathcal{H}$ to satisfy

$$|a'_{k_0}(t)|^{-1} \prec L(t) \prec |a''_{k_0}(t)|^{-1/2}$$

and

$$L(t) \prec (\psi'(t))^{-1}$$

for all the functions ψ of the set $\mathcal{A} = \{a_1 - a_{k_0}, ..., a_{k_0-1} - a_{k_0}, a_{k_0+1}, ..., a_k\}$. To see that such a function exists, we only need to prove that for any function $\psi \in \mathcal{A}$, we have

$$(a'_{k_0}(t))^{-1} \prec (\psi'(t))^{-1}$$

and

$$(a'_{k_0}(t))^{-1} \prec |a''_{k_0}(t)|^{-1/2}.$$

The first relation follows easily from the fact that a_{k_0} dominates all functions in $\mathcal{A}$ and L' Hospital's rule. The second relation also follows from L' Hospital's rule, since $\log t \prec a_{k_0}(t) \prec t$.

Using similar approximations as in the proof of Proposition 3.6.1, we deduce that for r sufficiently large, we can write

$$\lfloor \psi(n) \rfloor = \lfloor \psi(r) \rfloor + e_{\psi,n} \quad \text{for} \quad n \in [r, r + L(r)]$$

for every $\psi \in \mathcal{A}$, where $e_{\psi,n} \in \{0, \pm 1\}$. In addition, we can write

$$\lfloor a_{k_0}(n) \rfloor = \big\lfloor a_{k_0}(r) + (n-r)a'_{k_0}(r) \big\rfloor + e_{a_{k_0},n} \quad \text{for} \quad n \in [r, r + L(r)],$$

where $e_{a_{k_0},n} \in \{0, \pm 1\}$. Using the argument Lemma 3.2.1 once more to remove the error terms, our original problem reduces to showing

$$\limsup_{R\to+\infty} \sup_{\|f_2\|_\infty,...,\|f_k\|_\infty\leq 1} \underset{1\leq r\leq R}{\mathbb{E}} \sup_{|c_{h,r}|\leq 1} \Big\| \underset{0\leq h\leq L(r)}{\mathbb{E}} c_{h,r} T^{\lfloor a_{k_0}(r)+ha'_{k_0}(r)\rfloor} \Big(\prod_{i=1}^{k_0} T^{\lfloor (a_i(r)-a_{k_0}(r))\rfloor} f_i\Big) \prod_{i=k_0+1}^{k} T^{\lfloor a_i(r)\rfloor} f_i\Big\|_{L^2(\mu)} \ll_k \|\!|f_1|\!\|_{2k}. \quad (3.31)$$

Since the functions f_i are bounded by 1, the last relation follows if we prove that

$$\limsup_{R\to+\infty} \sup_{\|f_2\|_\infty,...,\|f_k\|_\infty\leq 1} \underset{1\leq r\leq R}{\mathbb{E}} \sup_{|c_{h,r}|\leq 1} \Big\| \underset{0\leq h\leq L(r)}{\mathbb{E}} c_{h,r} T^{\lfloor a_{k_0}(r)+ha'_{k_0}(r)\rfloor} \Big(\prod_{i=1}^{k_0} T^{\lfloor (a_i(r)-a_{k_0}(r))\rfloor} f_i\Big)\Big\|_{L^2(\mu)} \ll_k \|\!|f_1|\!\|_{2k}.$$

We choose functions $f_{2,R}, ..., f_{k_0,R}$ so that the corresponding average is $1/R$ close to the supremum. Write $F_{r,R} := \prod_{i=1}^{k_0} T^{\lfloor (a_i(r)-a_{k_0}(r))\rfloor} f_{i,R}$. We also fix a positive integer M. Repeating the same argument as in the base case, we can show that

$$\sup_{|c_{h,r}|\leq 1} \Big\| \mathop{\mathbb{E}}_{0\leq h\leq L(r)} c_{h,r}\ T^{\left\lfloor a_{k_0}(r)+ha'_{k_0}(r)\right\rfloor} F_{r,R} \Big\|^2_{L^2(\mu)} \ll \\ \frac{1}{M} + \sum_{z\in\{0,\pm1\}} \mathop{\mathbb{E}}_{|m|\leq M} \Big| \int F_{r,R}\ \cdot T^{m+z} F_{r,R}\ d\mu \Big| + o_r(1). \quad (3.32)$$

Therefore, we have

$$\sup_{\|f_2\|_\infty,\dots,\|f_k\|_\infty\leq 1} \mathop{\mathbb{E}}_{1\leq r\leq R} \sup_{|c_{h,r}|\leq 1} \Big\| \mathop{\mathbb{E}}_{0\leq h\leq L(r)} c_{h,r}\ T^{\left\lfloor a_{k_0}(r)+ha'_{k_0}(r)\right\rfloor} \big(\prod_{i=1}^{k_0} T^{\left\lfloor (a_i(r)-a_{k_0}(r))\right\rfloor} f_i\big) \Big\|^2_{L^2(\mu)} \ll \\ \frac{1}{M} + \mathop{\mathbb{E}}_{1\leq r\leq R} \sum_{z\in\{0,\pm1\}} \mathop{\mathbb{E}}_{|m|\leq M} \Big| \int F_{r,R} \cdot T^{m+z} F_{r,R}\ d\mu \Big| + O_R(1) \quad (3.33)$$

and we want to bound this last quantity by $O_k(1)$ times $\|\!|f_1|\!\|^2_{2k}$.

For a fixed $m\in[-M,M]$ and $z\in\{0,\pm1\}$, we apply the Cauchy-Schwarz inequality to get

$$\mathop{\mathbb{E}}_{1\leq r\leq R} \Big| \int F_{r,R}\cdot T^{m+z} F_{r,R}\ d\mu \Big| \leq \Big(\mathop{\mathbb{E}}_{1\leq r\leq R} \Big| \int F_{r,R}\cdot T^{m+z} F_{r,R}\ d\mu \Big|^2 \Big)^{1/2} = \\ \Big(\int \mathop{\mathbb{E}}_{1\leq r\leq R} (\overline{F_{r,R}}\otimes F_{r,R})\ (T\times T)^{m+z} (F_{r,R}\otimes \overline{F_{r,R}})\ d(\mu\times\mu) \Big)^{1/2} = \\ \Big(\int \mathop{\mathbb{E}}_{1\leq r\leq R} \prod_{i=1}^{k_0} (T\times T)^{\left\lfloor a_i(r)-a_{k_0}(r)\right\rfloor} \big((\overline{f_{i,R}}\otimes f_{i,R})\cdot (T\times T)^{m+z}(f_{i,R}\otimes \overline{f_{i,R}}) \big)\ d(\mu\times\mu) \Big)^{1/2} \leq \\ \Big\| \mathop{\mathbb{E}}_{1\leq r\leq R} \prod_{i=1}^{k_0-1} (T\times T)^{\left\lfloor a_i(r)-a_{k_0}(r)\right\rfloor} \big((\overline{f_{i,R}}\otimes f_{i,R})\cdot (T\times T)^{m+z}(f_{i,R}\otimes \overline{f_{i,R}}) \big) \Big\|^{1/2}_{L^2(\mu\times\mu)}$$

where $f_{1,R}=f_1$. Note that the functions $a_1-a_{k_0},\dots,a_{k-1}-a_{k_0}$ satisfy the hypotheses of Proposition 3.6.3. Therefore, we can apply the induction hypothesis (for $k_0-1<k$) to conclude that

$$\Big\| \mathop{\mathbb{E}}_{1\leq r\leq R} \prod_{i=1}^{k_0-1} (T\times T)^{\left\lfloor a_i(r)-a_{k_0}(r)\right\rfloor} \big((\overline{f_{i,R}}\otimes f_{i,R})\cdot (T\times T)^{m+z}(f_{i,R}\otimes \overline{f_{i,R}}) \big) \Big\|^{1/2}_{L^2(\mu\times\mu)} \ll_{k_0} \\ \|\!| (\overline{f_1}\otimes f_1)\cdot (T\times T)^{m+z}(f_1\otimes\overline{f_1}) |\!\|^{1/2}_{2k_0-2,T\times T}$$

and the last quantity is smaller than $\|\!|\overline{f_1}\cdot T^{m+z} f_1|\!\|_{2k_0-1,T}$. Putting this in (3.33), we get

$$\sup_{\|f_2\|_\infty,\dots,\|f_k\|_\infty\leq 1} \mathop{\mathbb{E}}_{1\leq r\leq R} \sup_{|c_{h,r}|\leq 1} \Big\| \mathop{\mathbb{E}}_{0\leq h\leq L(r)} c_{h,r}\ T^{\left\lfloor a_{k_0}(r)+ha'_{k_0}(r)\right\rfloor} \big(\prod_{i=1}^{k_0} T^{\left\lfloor (a_i(r)-a_{k_0}(r))\right\rfloor} f_i\big) \Big\|^2_{L^2(\mu)} \ll_k \\ \frac{1}{M} + \sum_{z\in\{0,\pm1\}} \mathop{\mathbb{E}}_{|m|\leq M} \|\!|\overline{f_1}\cdot T^{m+z} f_1|\!\|_{2k_0-1,T} + o_R(1) \leq \\ \frac{1}{M} + \sum_{z\in\{0,\pm1\}} \mathop{\mathbb{E}}_{|m|\leq M} \|\!|\overline{f_1}\cdot T^{m+z} f_1|\!\|_{2k-1,T} + o_R(1),$$

since $k_0\leq k$. Taking $R\to+\infty$ and then $M\to+\infty$, we get that it suffices to show that

$$\limsup_{M\to+\infty} \mathop{\mathbb{E}}_{|m|\leq M} \|\!|\overline{f_1}\cdot T^{m+z} f_1|\!\|_{2k-1} \leq \|\!|f_1|\!\|^2_{2k}$$

for any $z\in\{0,\pm1\}$. This follows easily by raising to the 2^{2k-1}-th power and using the Hölder inequality, as well as the definition of the Host-Kra seminorms. □

3.7 The general case of Proposition 3.1.1

In this section we aim to prove main proposition of this chapter. We maintain the notation of Proposition 3.1.1 and we also assume that at least on of the functions $a_1, ..., a_k$ has super-linear growth. We also consider the set of functions

$$S = \{a_1(t), a_1(t) - a_2(t), ..., a_1(t) - a_k(t)\}$$

Functions in S dominate $\log t$ by our hypothesis. Finally, we assume that not every one of the involved functions has the form $p(t) + g(t)$, where $p \in \mathbb{R}[t]$ and $g \in \mathcal{H}$ is sub-fractional, since this case was covered in the previous section (it follows from Corollary 3.6.2). In particular, we assume that this holds for the function a_1.

We will use the following decomposition result from [49].

Lemma 3.7.1. *[49, Lemma A.3] Let $a_1, ..., a_k \in \mathcal{H}$ have polynomial growth. Then, there exist a natural number m, functions $g_1, ..., g_m \in \mathcal{H}$, real numbers $c_{i,j}$, where $1 \leq i \leq k$ and $1 \leq j \leq m$, and real polynomials $p_1, ..., p_k$ such that:*

1. *$g_1 \prec g_2 \prec ... \prec g_m$,*
2. *$t^{l_i} \prec g_i(t) \prec t^{l_i+1}$ for some $l_i \in \mathbb{Z}^+$ (i.e. they are strongly non-polynomial) and*
3. *for all $i \in \{1, 2, ..., k\}$ we have*

$$a_i(t) = \sum_{j=1}^{m} c_{i,j} g_j(t) + p_i(t) + o_t(1).$$

Note that the functions g_j do not necessarily belong in the set of linear combinations of the $a_1, ..., a_k$. The proof of this lemma can be found in the appendix of [49]. As an example, if we have the pair $\{t+t^{3/2}, t^2+t^{5/2}\}$, then the functions in the above decomposition are $\{g_1, g_2, p_1, p_2\} = \{t^{3/2}, t^{5/2}, t, t^2\}$.

Returning to our original problem, we split the given family of functions into two sets

$$J_1 = \{a_i \colon a_i(t) \ll t^{\delta} \text{ for all } \delta > 0\} \text{ and } J_2 = \{a_i \colon \exists\, \delta > 0 \text{ with } \ a_i(t) \gg t^{\delta}\}.$$

We do the same for the set S of differences:

$$S_1 = \{f \in S \colon f(t) \ll t^{\delta} \text{ for all } \delta > 0\} \text{ and } S_2 = \{f \in S \colon \exists\, \delta > 0 \text{ with } \ f(t) \gg t^{\delta}\}.$$

Observe that the function a_1 belongs to the sets J_2 and S_2 due to our assumption in the beginning of this section.

We will see that the slow-growing functions in sets J_1 and S_1 will be approximately equal to a constant, when we consider averages on small intervals. For the remaining functions, we will use the Taylor expansion to approximate them. We split the proof into several steps. Steps 1 through 4 of this proof correspond to step 1 in example a) of section 3.1, while steps 5 and 6 of the proof correspond to step 2 of the same example. The remaining two steps correspond to step 3 of example a). In Step 8, we will also use the results of the special case of the previous section.

3.7.1 Step 1: Introducing a double averaging

Let $L(t) \in \mathcal{H}$ be a sub-linear function to be specified later. We can consider a priori functions that satisfy $L(t) \prec t^{1-\varepsilon}$ for some $\varepsilon > 0$ (i.e. we exclude functions like $t/\log t$). Invoking Lemma 3.2.2, we see that it is sufficient to prove that

$$\limsup_{R\to\infty} \underset{1\leq r\leq R}{\mathbb{E}} \sup_{|c_{r,n}|\leq 1} \Big\| \underset{r\leq n\leq r+L(r)}{\mathbb{E}} c_{r,n}\, T^{\lfloor a_1(n)\rfloor} f_1 \cdot ... \cdot T^{\lfloor a_k(n)\rfloor} f_{k,R} \Big\|_{L^2(\mu)}^{2^t} = 0 \tag{3.34}$$

for any sequences of 1-bounded functions $f_{2,R}, ..., f_{k,R}$ and some positive integer parameter t, which will depend only on the original functions $a_1, ..., a_k$. Therefore, when applying Lemma 3.2.1, we can always assume that the implicit constant (which depends on the exponent 2^t) is an $O_{a_1,...,a_k}(1)$ constant.

We observe that (3.34) follows if we show that

$$\sup_{||f_2||_\infty\le 1,\dots,||f_k||_\infty\le 1} \mathop{\mathbb{E}}_{1\le r\le R} \sup_{|c_{r,n}|\le 1} \Big\| \mathop{\mathbb{E}}_{r\le n\le r+L(r)} c_{r,n}\, T^{\lfloor a_1(n)\rfloor} f_1\cdot \ldots \cdot T^{\lfloor a_k(n)\rfloor} f_k \Big\|_{L^2(\mu)}^{2^t} \tag{3.35}$$

goes to 0, as $R\to+\infty$.

3.7.2 Step 2: Eliminating the small functions of J_1

While in example a) of Section 4 we used the Taylor expansion right at the beginning, it is more convenient to reverse our steps a bit in the proof.

Assume that the function a_i belongs to the set J_1 (namely, it is a sub-fractional function). Then, for any $n\in[r,r+L(r)]$, we have

$$|a_i(n)-a_i(r)|=|n-r||a_i'(\xi)|$$

for some $\xi\in[r,n]$. Since $|a_i'(t)|\searrow 0$, we get

$$|a_i(n)-a_i(r)|\le |L(r)||a_i'(r)|,$$

which is $o_r(1)$. Note that we already assumed that we will eventually choose $L\in\mathcal{H}$ such that $L(t)\ll t^{1-\varepsilon}$, which makes the previous statements valid (see the discussion at the end of the Appendix). Thus, if r is sufficiently large and $n\in[r,r+L(r)]$, we can write $\lfloor a(n)\rfloor=\lfloor a(r)\rfloor+\varepsilon_{r,n}$, where $\varepsilon_{r,n}\in\{0,1\}$. Using the argument in Lemma 3.2.1, we absorb the error terms $\varepsilon_{r,n}$ in the supremum outside of the averages in (3.35).

The iterate corresponding to the function f_i has now become constant and we can ignore it. In conclusion, we have reduced our problem to the case that the set J_1 is empty.

3.7.3 Step 3: Concatenating the functions of the set S_1

Assume that the function a_1-a_i belongs to S_1. Then, mimicking the arguments of the previous step, we can write $a_i=a_1+(a_i-a_1)$ where the function a_i-a_1 is asymptotically a constant in the interval $[r,r+L(r)]$. Then, we can combine the product of all such terms

$$T^{\lfloor a_1(n)\rfloor} f_1 \prod_{a_1-a_i\in S_1} T^{\lfloor a_i(n)\rfloor} f_i$$

into one iterate $T^{\lfloor a_1(n)\rfloor}\tilde{f}_r$ (we use again the argument in Lemma 3.2.1 to remove the error terms), where

$$\tilde{f}_r=f_1\cdot T^{\lfloor \theta_1(r)\rfloor}h_1\cdot\ldots\cdot T^{\lfloor\theta_\ell(r)\rfloor}h_\ell, \tag{3.36}$$

where $h_1,\dots,h_\ell$ are functions in $L^\infty(\mu)$ and the functions $\theta_1,\dots,\theta_\ell\in\mathcal{H}$ are sub-linear functions that satisfy

$$\log t\prec\theta_i(t)\prec t^\delta$$

for all $\delta>0$. In addition, the assumption that the pairwise differences of the functions $a_1,\dots,a_k$ dominate $\log t$ implies that

$$\log t\prec\theta_i(t)-\theta_j(t)$$

for $i\neq j$.

Now the original problem reduces to the following: If all the functions $a_1,\dots,a_k$ are such that the sets J_1 and S_1 are empty, then show that the averages

$$\sup_{||f_2||_\infty,\dots,||f_k||_\infty\le 1}\ \sup_{||h_1||_\infty,\dots,||h_\ell||_\infty\le 1} \mathop{\mathbb{E}}_{1\le r\le R} \sup_{|c_{r,n}|\le 1} \Big\| \mathop{\mathbb{E}}_{r\le n\le r+L(r)} c_{r,n}\, T^{\lfloor a_1(n)\rfloor}\tilde{f}_r\cdot\ldots\cdot T^{\lfloor a_k(n)\rfloor} f_k\Big\|_{L^2(\mu)}^{2^t} \tag{3.37}$$

go to 0 as $R\to+\infty$, where the function $\tilde{f}_r$ is the function in (3.36).

We can repeat the same argument of this step to reduce to the case where $a_i(t)-a_j(t)\gg t^\delta$ for some $\delta>0$. Indeed, if the difference a_i-a_j is sub-fractional, we can combine the iterates corresponding to these two functions into a single iterate of the form $T^{\lfloor a_i(n)\rfloor}g_r$ for some function g_r. In order to replace g_r by a function that does not depend on r, we move the supremum of the $f_2,...,f_k$ inside the outer average. In conclusion, it suffices to show that

$$\sup_{||h_1||_\infty,\dots,||h_\ell||_\infty\le 1}\ \mathop{\mathbb{E}}_{1\le r\le R}\ \sup_{||f_2||_\infty,\dots,||f_k||_\infty\le 1}\ \sup_{|c_{r,n}|\le 1}\Big\|\mathop{\mathbb{E}}_{r\le n\le r+L(r)} c_{r,n}\ T^{\lfloor a_1(n)\rfloor}\tilde{f}_r\cdot\ldots\cdot T^{\lfloor a_k(n)\rfloor}f_k\Big\|_{L^2(\mu)}^{2^t} \tag{3.38}$$

goes to 0 as $R\to+\infty$, where $\tilde{f}_r$ is the function in (3.36) and all differences a_i-a_j dominate some fractional power[9]. Recall that the functions θ_i satisfy

$$\log t\prec\theta_i(t)\prec t^\delta\ \text{ for every }\ \delta>0$$

and

$$\log t\prec\theta_i(t)-\theta_j(t).$$

3.7.4 Step 4: Approximating by polynomials

In this step, we will use the Taylor expansion to replace the functions a_i by polynomials in the intervals $[r,r+L(r)]$. First of all, we can use Lemma 3.7.1 in order to write

$$a_i(t)=\sum_{j=1}^m c_{i,j}g_j(t)+q_i(t)+o_t(1), \tag{3.39}$$

where $g_1\prec g_2\prec\ldots\prec g_m$ are strongly non-polynomial functions and $q_i(t)$ are real polynomials. We immediately conclude that the function g_m cannot be sub-fractional. Indeed, if that was the case, then all the functions a_i would be a sum of a polynomial plus a sub-fractional function, which is at odds with our initial assumption.

The $o_t(1)$ terms can be eliminated by using an argument similar to the proof of Lemma 3.2.1. In addition, we may assume that $c_{1,m}\neq 0$ (and thus g_m exists in the expansion of a_1). This can be proven by an argument similar to the one in the beginning of Section 3.1 (the same reasoning we used to reduce our problem to the case that a_1 has maximal growth rate). Of course, by assuming this new property, we abandon the assumption that a_1 has maximal growth rate.

We define

$$\mathcal{F}=\{g_1,...,g_m\}$$

and let $\mathcal{A}=\{g_1,...,g_l\}\subseteq\mathcal{F}$ be the set of functions that satisfy $g_i(t)\ll t^\delta$ for all $\delta>0$ (i.e the sub-fractional functions). We have that $g_m\notin\mathcal{A}$.

By the reductions in steps 2 and 3, we have that $a_i(t)\gg t^{\delta_i}$ for some $\delta_i>0$ and a similar relation holds for the differences a_i-a_j. Therefore, we have the following property:

$$\text{If } i_1\neq i_2,\ \text{we have either } c_{i_1,j}\neq c_{i_2,j}\ \text{ for some } j>l,\ \text{or } q_{i_1}(t)-q_{i_2}(t) \text{ is non-constant.} \tag{\textbf{P}}$$

Now every function $g\in\mathcal{A}$ satisfies

$$\max_{n\in[r,r+L(r)]}|g(n)-g(r)|=o_r(1)$$

by the arguments in the preceding steps. We can use the argument in Lemma 3.2.1 to remove the error term $o_r(1)$ and then substitute each function $g\in\mathcal{A}$ in the interval $[r,r+L(r)]$ by a constant (namely, the value of the function g at r). These constants can be absorbed by the supremum of the

[9] Since our functions $a_1,...,a_k$ dominate a fractional power, we can now use the fact that the classes $S(a_i,k)$ (defined and studied in the Appendix) can be well defined in order to approximate all of them by polynomials.

$f_2, ..., f_k$ and the use of Lemma 3.2.1. Therefore, we may assume that all functions $g_1, ..., g_m$ dominate some fractional power t^δ (equivalently $\mathcal{A} = \emptyset$) and that property (**P**) above holds with $l = 0$.

Since the functions $g_1, ..., g_m$ dominate some fractional power, the classes

$$S(g_i, n) = \{f \in \mathcal{H}, (g_i^{(n)}(t))^{-1/n} \preceq f(t) \prec (g_i^{(n+1)}(t))^{-1/(n+1)}\}$$

are well defined for n large enough. We remind the reader that these classes and their properties are all studied in the Appendix and we will use them freely from this point onward.

Let d be a natural number and for every function $g \in \mathcal{F}$, we consider the natural number k_g, such that the function $|g_m^{(d)}(t)|^{-\frac{1}{d}}$ belongs to the class $S(g, k_g)$. This class always exists, if we pick our number d to be sufficiently large. We immediately deduce that $k_g \leq d$ for every $g \in \mathcal{F}$, while $k_{g_m} = d$.

Let q be a positive real number (but not an integer), such that t^q dominates all functions $g_1, ..., g_m$ and the polynomials $q_1, ..., q_k$. In particular, this implies that, for all $1 \leq i \leq m$, all derivatives of g_i of order bigger than q go to 0 (as $t \to +\infty$). This is a consequence of the growth inequalities of 2.1.3. We make the additional assumption that our integers k_g are very large compared to q, which can be attained if we take our initial number d to be sufficiently large. The inequality $k_g \geq 10q$ will suffice for our purposes.

Definition 3.7.2. *We say that two functions $f \ll g$ of $\mathcal{H}$ have the property $\mathcal{Q}$, if they have the same growth rate, or if $g(t)$ strongly dominates $f(t)$.*

We remind the reader that we say that $g(t)$ strongly dominates $f(t)$ and write $g(t) \ggg f(t)$, if the ratio

$$\frac{g(t)}{f(t)}$$

dominates some fractional power t^δ, $\delta > 0$ (see also our notational conventions in the first chapter).

We consider two possible cases:

a) Assume that for every $g \in \mathcal{F} \setminus \{g_m\}$, the functions $|g_m^{(d)}(t)|^{-\frac{1}{d}}$ and $|g^{(k_g)}(t)|^{-\frac{1}{k_g}}$ have the property[10] $\mathcal{Q}$. Then, our selection will be the classes $S(g, k_g)$ as they stand. Furthermore, we choose $L(t) \in \mathcal{H}$ to be any function that belongs to the intersection of the classes $S(g_i, k_{g_i})$ (which is non-empty by definition). In this case, we call the function g_m our "special" function. Note that

$$|g^{(k_g)}(t)|^{-\frac{1}{k_g}} \preceq |g_m^{(d)}(t)|^{-\frac{1}{d}}$$

for $g \neq g_m$ in this case.

b) Assume that the above case does not hold[11]. Then, among all the functions $|g^{(k_g)}(t)|^{-\frac{1}{k_g}}$ for which the property $\mathcal{Q}$ fails (in relation to $|g_m^{(d)}(t)|^{-\frac{1}{d}}$), we choose a function g for which $|g^{(k_g)}(t)|^{-\frac{1}{k_g}}$ has minimal growth rate. Then, we choose a function $L(t) \in \mathcal{H}$ with the following properties:

i) If a function $\tilde{g}$ is such, that $|\tilde{g}^{(k_{\tilde{g}})}(t)|^{-\frac{1}{k_{\tilde{g}}}}$ fails to satisfy property $\mathcal{Q}$ in relation to $|g_m^{(d)}(t)|^{-\frac{1}{d}}$ and has different growth rate than g, then we have

$$|(\tilde{g})^{(k_{\tilde{g}}-1)}(t)|^{-\frac{1}{k_{\tilde{g}}-1}} \prec L(t) \prec |(\tilde{g})^{(k_{\tilde{g}})}(t)|^{-\frac{1}{k_{\tilde{g}}}}.$$

Namely, we have $L(t) \in S(\tilde{g}, k_{\tilde{g}} - 1)$.

ii) If the function $\tilde{g}$ has the same growth rate as g, then we have $k_g = k_{\tilde{g}}$ and the classes $S(g, k_g)$ and $S(\tilde{g}, k_{\tilde{g}})$ coincide. In this case, we leave the integer $k_{\tilde{g}}$ as is and we will have $L(t) \in S(\tilde{g}, k_{\tilde{g}})$.

iii) The third case is when the function $\tilde{g}$ satisfies property $\mathcal{Q}$ in relation to $|g_m^{(d)}(t)|^{-\frac{1}{d}}$. Then, we leave the the integer $S(\tilde{g}, k_{\tilde{g}})$ as is and take $L(t) \in S(\tilde{g}, k_{\tilde{g}})$.

[10] An example of functions that fall in this case is the pair $(t^{3/2}, t \log t)$, if we consider their second derivatives. We can easily check that the ratio of the second derivatives of these two functions raised to the $-\frac{1}{2}$-th power grows like the function $t^{1/4}$.

[11] An example of functions that fall in this second case is the pair $(t \log t, t \log \log t)$, if we again consider their second derivatives. A simple computation yields that the growth rate of the ratio of the involved functions grows like the function $\sqrt{\log t}$ and, thus, they fail property $\mathcal{Q}$.

The existence of such a function $L(t)$ follows by our minimality assumption on $|g^{(k_g)}(t)|^{-\frac{1}{k_g}}$. In this case, g is our "special" function.

We denote by k'_g the new integers that appear after the above procedure.

Claim 2. *For the choice we have made above, the function* $|z^{(k'_z)}(t)|^{-\frac{1}{k'_z}}$ *satisfies property* $(\mathcal{Q})$ *in relation to our special function, for any* $z \in \mathcal{F}$.

Proof. If we are in case a) above, the functions $|g_m^{(d)}(t)|^{-\frac{1}{d}}$ and $|g^{(k_g)}(t)|^{-\frac{1}{k_g}}$ have the same growth rate or their ratio dominates a fractional power (for any $g \in \mathcal{F}$) and we are done.

In case b), we have a special function g $(k_g = k'_g)$. We consider functions $z \neq g$ such that $|g^{(k_g)}(t)|^{-\frac{1}{k_g}}$ and $|z^{(k'_z)}(t)|^{-\frac{1}{k'_z}}$ have different growth rates (because otherwise the claim is trivial). Then there are two possibilities:

• If the original function $|g_m^{(d)}(t)|^{-\frac{1}{d}}$ and $|z^{(k_z)}(t)|^{-\frac{1}{k_z}}$ had a ratio dominating a fractional power, then the claim follows (in this case, we must have $k'_z = k_z$).

• If the original function $|z^{(k_z)}(t)|^{-\frac{1}{k_z}}$ failed property $\mathcal{Q}$ in relation to $|g_m^{(d)}(t)|^{-\frac{1}{d}}$, then we have

$$|g^{(k_g)}(t)|^{-\frac{1}{k_g}} \prec |z^{(k_z)}(t)|^{-\frac{1}{k_z}} \quad \text{(due to minimality)}$$

and thus $L(t) \in S(z, k_z - 1)$. We easily see that the functions $|g^{(k_g)}(t)|^{-\frac{1}{k_g}}$ and $|z^{(k_z-1)}(t)|^{-\frac{1}{k_z-1}}$ differ by a fractional power. Indeed, we have a "gain" of some power t^{δ} when passing from $S(z, k_z - 1)$ to $S(z, k_z)$ due to (3.21). Therefore, if the functions $|z^{(k_z)}(t)|^{-\frac{1}{k_z}}$ and $|g^{(k_g)}(t)|^{-\frac{1}{k_g}}$ were "close", then $|z^{(k_z-1)}(t)|^{-\frac{1}{k_z-1}}$ and $|g^{(k_g)}(t)|^{-\frac{1}{k_g}}$ differ by a fractional power. □

For convenience, we will use the same notation $S(g, k_g)$ for the new classes that have been chosen after the above operation (that is we replace k'_g by k_g).

Remark. The above proof also implies that the growth rate of $|g^{(k_g)}(t)|^{-\frac{1}{k_g}}$ is maximized when g is the special function.

We denote by $\tilde{g}$ the special function given by our above arguments. For any function $g \in \mathcal{F}$, we use the Taylor expansion around the point r to obtain

$$g(r+h) = g(r) + \cdots + \frac{g^{(k_g)}(r)h^{k_g}}{k_g!} + \frac{g^{(k_g+1)}(\xi_m)h^{k_g+1}}{(k_g+1)!} \text{ for some } \xi_m \in [r, r+m], \tag{3.40}$$

for all $0 \leq h \leq L(r)$. We observe that the last term is $o_r(1)$ while the second to last term in the above expansion diverges when $h = L(r)$. Therefore, we have

$$g(r+h) = p_{r,g}(h) + o_r(1)$$

where $p_{r,g}$ is a polynomial.

3.7.5 Step 5: The change of variables

In this step, we do a change of variables trick. Our purpose is to rewrite the above polynomials in such a way, that the leading coefficients are good sequences in order to be able to apply Proposition 3.4.5. All the work we did in the previous step (namely, making sure that our functions satisfied Property $\mathcal{Q}$) will ensure that the leading coefficients of our polynomials will be good sequences that either converge to a (non-zero) real number, or their growth rate is larger than some fractional power. A similar trick is also used in [13], although the fact that the argument in this article involves essentially only one Hardy function makes the argument much simpler.

Assume that $\tilde{g}$ is our special function with the polynomial expansion

$$\tilde{g}(r+h) = \tilde{g}(r) + \cdots + \frac{\tilde{g}^{(k_{\tilde{g}})}(r)h^{k_{\tilde{g}}}}{k_{\tilde{g}}!} + o_r(1).$$

Every $0 \le h \le L(r)$ can be written as

$$h = w \lfloor \Big| \frac{k_{\tilde{g}}!}{\tilde{g}^{(k_{\tilde{g}})}(r)} \Big|^{\frac{1}{k_{\tilde{g}}}} \rfloor + v$$

for some integers w, v, where

$$0 \le w \le \frac{L(r)}{\lfloor \Big| \frac{k_{\tilde{g}}!}{\tilde{g}^{(k_{\tilde{g}})}(r)} \Big|^{\frac{1}{k_{\tilde{g}}}} \rfloor} = D_r$$

and

$$0 \le v \le \lfloor \Big| \frac{k_{\tilde{g}}!}{\tilde{g}^{(k_{\tilde{g}})}(r)} \Big|^{\frac{1}{k_{\tilde{g}}}} \rfloor - 1.$$

Note that $D_r \succ 1$, because $L(t) \in S(\tilde{g}, k_{\tilde{g}})$. We denote by $u(r)$ the function inside the integer part above, namely, we define

$$u(t) = \Big| \frac{k_{\tilde{g}}!}{\tilde{g}^{(k_{\tilde{g}})}(t)} \Big|^{\frac{1}{k_{\tilde{g}}}},$$

which is a (sub-linear) function in $\mathcal{H}$. In addition, since we have chosen the numbers k_g to be sufficiently large, we can ensure that the function u dominates some fractional power (this follows by statement ii) of Lemma 3.5.1).

We observe that (recall that $\tilde{f}_r$ is given by (3.36))

$$\sup_{\|f_2\|_\infty, \dots, \|f_k\|_\infty \le 1} \Big\| \mathop{\mathbb{E}}_{0 \le h \le L(r)} c_{h,r} \, T^{\lfloor a_1(r+h) \rfloor} \tilde{f}_r \cdot \ldots \cdot T^{\lfloor a_k(r+h) \rfloor} f_k \Big\|_{L^2(\mu)}^{2^t} \le$$
$$\sup_{\|f_2\|_\infty, \dots, \|f_k\|_\infty \le 1} \mathop{\mathbb{E}}_{1 \le v \le \lfloor u(r) \rfloor - 1} \sup_{|c_{h,r,v}| \le 1} \Big\| \mathop{\mathbb{E}}_{h \equiv v (mod \, \lfloor u(r) \rfloor)} c_{h,r,v} \, T^{\lfloor a_1(r+h) \rfloor} \tilde{f}_r \cdot \ldots \cdot T^{\lfloor a_k(r+h) \rfloor} f_k \Big\|_{L^2(\mu)}^{2^t}, \tag{3.41}$$

where the above bound follows by applying the Hölder and triangle inequalities. We will bound the innermost average in the norm by a quantity that does not depend on v.

Fix a v as above. For every $h \equiv v (mod \, \lfloor u(r) \rfloor)$, we can write each of the polynomials $p_{g,r}(h)$ in the previous step as a new polynomial $\tilde{p}_{r,v,g}(w)$ in the new variable w. We are only interested in the leading coefficients of the new polynomials. Using (3.40), we see that it is equal to

$$c_g(r) = \frac{g^{(k_g)}(r)}{k_g!} \cdot \lfloor u(r) \rfloor^{k_g} = \frac{g^{(k_g)}(r)}{k_g!} \cdot \lfloor \Big| \frac{k_{\tilde{g}}!}{\tilde{g}^{(k_{\tilde{g}})}(r)} \Big|^{\frac{1}{k_{\tilde{g}}}} \rfloor^{k_g}. \tag{3.42}$$

Now assume that $g \in \mathcal{F}$. The function $c_g(r)$ is not a function in the Hardy field $\mathcal{H}$, but we will prove that it is a good sequence (see Definition 3.4.1). Therefore, we seek to approximate it by a function in $\mathcal{H}$. The simplest way to achieve this is to define the function $d_g(t) \in \mathcal{H}$ by removing the floor function:

$$d_g(t) = \frac{g^{(k_g)}(t)}{k_g!} \cdot \Big| \frac{k_{\tilde{g}}!}{\tilde{g}^{(k_{\tilde{g}})}(t)} \Big|^{\frac{k_g}{k_{\tilde{g}}}}. \tag{3.43}$$

It is obvious that $c_g(r)/d_g(r) \to 1$. However, we have something stronger:

Claim 3. *For all $g \in \mathcal{F}$, we have*

$$|c_g(r) - d_g(r)| = o_r(1).$$

Proof. We will use the inequality

$$|a^c - b^c| \le c|a-b||a|^{c-1},$$

which holds when $|b| \le |a|$ and $c \in \mathbb{N}$. An application of this inequality reduces the problem to showing that

$$|\tilde{g}^{(k_{\tilde{g}})}(t)|^{-\frac{1}{k_{\tilde{g}}}} \prec |g^{(k_g)}(t)|^{-\frac{1}{k_g - 1}}. \tag{3.44}$$

Since $L(t) \in S(\tilde{g}, k_{\tilde{g}})$, it is sufficient to show that

$$L(t) \prec |g^{(k_g)}(t)|^{-\frac{1}{k_g-1}}$$

and now using the fact that $L(t) \in S(g, k_g)$, our conclusion follows if we prove that

$$|g^{(k_g+1)}(t)|^{-\frac{1}{k_g+1}} \prec |g^{(k_g)}(t)|^{-\frac{1}{k_g-1}}.$$

Using the inequality $g^{(k_g+1)}(t) \succ g^{(k_g)}(t)/t\log^2 t$ in the above equation (we use Lemma 2.1.3 and the fact that the numbers k_g are assumed to be large enough), this reduces to

$$g^{(k_g)}(t) \prec t^{\frac{1-k_g}{2}} \log^{(1-k_g)} t. \tag{3.45}$$

However, recall that we have chosen a non-integer q, such that $g(t) \ll t^q$ for all $g \in \mathcal{F}$ and we have also chosen $k_g \geq 10q - 1$. Applying Lemma 2.1.3, we have $g^{(k_g)}(t) \prec t^{q-k_g}$ and now the claim easily follows. □

Claim 4. *a) We have that the function $d_g(t)$ in (3.43) is a sub-linear function that either satisfies $t^\varepsilon \prec d_g(t)$ for some $\varepsilon > 0$ or converges to a non-zero constant*[12]*.*
b) We have the growth relation $d_g(t) \prec (\tilde{g}^{(k_{\tilde{g}})}(t))^{-\frac{1}{k_{\tilde{g}}}}$ and, thus, d_g has sub-linear growth.

Proof. Property ($\mathcal{Q}$) implies that $d_g(t)$ converges to a non-zero constant, or dominates a fractional power t^δ. For the second part, we observe that a simple computation shows that this is equivalent to (3.44), which has already been established. □

Claim 5. *If g, h are distinct functions in the set $\{g_1, ..., g_m\}$ such that $d_g(t) \sim d_h(t)$, then $k_g \neq k_h$.*

Proof. Assume that we have both $k_g = k_h$ and $d_g \sim d_h$. This implies that

$$g^{(k_g)}(t) \sim h^{(k_h)}(t)$$

and L'Hospital' rule implies that $g \sim h$. Since g, h have distinct growth rates and are strongly non-polynomial, this last relation cannot hold and we arrive at a contradiction. □

We have seen that the functions $g_1, ..., g_m$ admit a polynomial expansion and, after the change of variables above, their leading coefficients become sub-linear good sequences. Now, we look how the leading coefficients of the polynomials $q_1, ..., q_k$ in (3.39) transform after the above change of variables. Note that $q_i(r+h)$ is also a polynomial $q_{i,r}(h)$ in the variable h. Writing again

$$h = w\lfloor u(r) \rfloor + v$$

as above, we see that $q_i(r+h) = q_{i,r,v}(w)$ where $q_{i,r,v}$ is a real polynomial. It is straightforward to check that the leading coefficients of the $q_{i,r,v}$ have the form $c\lfloor u(r) \rfloor^\theta$, where $c \in \mathbb{R}^*$ and $\theta \in \mathbb{N}^+$. These are good sequences, since they are asymptotically equal to

$$c\Big|\frac{k_{\tilde{g}}!}{\tilde{g}^{(k_{\tilde{g}})}(r)}\Big|^{\frac{\theta}{k_{\tilde{g}}}},$$

which is a function in $\mathcal{H}$ (and its limit is obviously non-zero).

Now, we recall (3.39). When restricted to the interval $[r, r + L(r)]$, every one of our original functions a_i, where $1 \leq i \leq k$ can be written as a sum of polynomials, whose leading coefficients are good sequences, plus an $o_r(1)$ term. We can eliminate the error terms $o_r(1)$ by using the argument in Lemma 3.2.1 once again. In particular, any one of these good sequences (denote a_r) satisfies one of the following:
a) there exists a sub-linear function $\phi \in \mathcal{H}$, such that $a_r = \phi(r) + o_r(1) \prec u(r)$ and $\phi(t) \gg t^\delta$ for some $\delta > 0$,

[12]Thus, the leading coefficients $c_g(r)$ in (3.42) are good sequences.

b) they have the form $c\lfloor u(r)\rfloor^{\theta}$, where $c\in\mathbb{R}$ and θ is a positive integer or
c) they converge to a non-zero real number.

We denote the polynomial corresponding to a_i as $P_{i,r,v}$ and we observe that its degree is independent of r. In view of Property (**P**), we deduce that the leading coefficient of $P_{i,r,v}-P_{j,r,v}$ is either the leading coefficient of the polynomial $q_{i,r,v}(t)-q_{j,r,v}(t)$ (which in this case must be a non-constant polynomial), or it is equal to the leading coefficient of

$$R_{ij.r.v}(w)=\sum_{n=1}^{m}(c_{i,n}-c_{j,n})\tilde{p}_{r,g_j,v}(w) \tag{3.46}$$

or it is a combination of these two coefficients. In the first case, it has the form b) above and is a good sequence. In the second case, it is a linear combination of sequences of the form *a*) or *c*). That is, there are functions $g_{i_1},...,g_{i_\lambda}$, where $i_1,...,i_\lambda\in\{1,2,...,m\}$ such that the leading coefficients of the polynomials $\tilde{p}_{r,g_{i_j},v}$ are all sequences of the form *a*) or *c*) and the leading coefficient of the polynomial $R_{ij,r,v}$ in (3.46) is equal to the leading coefficient of

$$\sum_{\alpha=1}^{\lambda}(c_{i,i_\alpha}-c_{j,i_\alpha})\tilde{p}_{r,g_{i_\alpha},v}. \tag{3.47}$$

We will use Claim 5: if any two of the polynomials $\tilde{p}_{r,g_{i_\alpha},v}$ have the same degree, then their leading coefficients are sequences with distinct growth rates. Therefore, the leading coefficient of $R_{ij,r,v}$ is a linear combination of good sequences with pairwise distinct growth rates and it is straightforward to see that it is itself a good sequence. Finally, we observe that the final case cannot happen (namely, a combination of these two coefficients). That is because the degree of the polynomial $q_{i,r,v}(t)-q_{j,r,v}(t)$, which is equal to the degree of q_i-q_j, is very small compared to the degree of the polynomial in (3.47), because we chose the degrees k_g of the polynomials in the Taylor expansions to be very large compared to the degrees of the polynomials $q_1,...,q_k$.

Our original problem reduces to the following (recall (3.41)): for every measure-preserving system (X,μ,T) and function $f_1\in L^\infty(\mu)$ with $f_1\perp Z_{\tilde{s}}(X)$ for some $\tilde{s}\in\mathbb{N}$, there exists a positive integer $t=t(a_1,...,a_k)$ such that:

$$\lim_{R\to+\infty}\sup_{||h_1||_\infty\leq 1,...,||h_\ell||_\infty\leq 1}\mathop{\mathbb{E}}_{1\leq r\leq R}\mathop{\mathbb{E}}_{0\leq v\leq\lfloor u(r)\rfloor-1}$$
$$\sup_{||f_2||_\infty\leq 1,...,||f_k||_\infty\leq 1}\sup_{|c_{w,r,v}|\leq 1}\Big\|\mathop{\mathbb{E}}_{0\leq w\leq D_r}c_{w,r,v}\,T^{\lfloor P_{1,r,v}(w)\rfloor}\tilde{f}_r\cdot...\cdot T^{\lfloor P_{k,r,v}(w)\rfloor}f_k\Big\|^{2^t}_{L^2(\mu)}=0, \tag{3.48}$$

where

$$\tilde{f}_r=f_1\cdot T^{\lfloor\theta_1(r)\rfloor}h_1\cdot...\cdot T^{\lfloor\theta_\ell(r)\rfloor}h_\ell \tag{3.49}$$

for functions $\theta_1,...,\theta_\ell\in\mathcal{H}$ that satisfy

$$\log t\prec\theta_i(t)\prec t^\delta$$
$$\log t\prec\theta_i(t)-\theta_j(t)\prec t^\delta\quad\text{for}\quad i\neq j$$

for all $\delta>0$.

Observe that

$$\mathop{\mathbb{E}}_{0\leq v\leq\lfloor u(r)\rfloor-1}\sup_{||f_2||_\infty\leq 1,...,||f_k||_\infty\leq 1}\sup_{|c_{w,r,v}|\leq 1}\Big\|\mathop{\mathbb{E}}_{0\leq w\leq D_r}c_{w,r,v}\,T^{\lfloor P_{1,r,v}(w)\rfloor}\tilde{f}_r\cdot...\cdot T^{\lfloor P_{k,r,v}(w)\rfloor}f_k\Big\|^{2^t}_{L^2(\mu)}\leq$$
$$\max_{0\leq v\leq\lfloor u(r)\rfloor-1}\sup_{||f_2||_\infty\leq 1,...,||f_k||_\infty\leq 1}\sup_{|c_{w,r,v}|\leq 1}\Big\|\mathop{\mathbb{E}}_{0\leq w\leq D_r}c_{w,r,v}\,T^{\lfloor P_{1,r,v}(w)\rfloor}\tilde{f}_r\cdot...\cdot T^{\lfloor P_{k,r,v}(w)\rfloor}f_k\Big\|^{2^t}_{L^2(\mu)}.$$

For each $r\in\mathbb{N}$, let v_r be the value of v for which the above max is attained. Then, the polynomial family

$$\mathcal{P}_r=\{P_{1,r,v_r},...,P_{k,r,v_r}\}$$

is a nice polynomial family. Indeed, the degrees of its elements are fixed integers and the leading coefficients of the polynomials and of their differences are good sequences irrespective of the value of v_r, as we discussed previously. Therefore, under the above assumptions, we reduce our problem to

$$\lim_{R\to+\infty} \sup_{||h_1||_\infty\leq 1,\ldots,||h_\ell||_\infty\leq 1} \mathop{\mathbb{E}}_{1\leq r\leq R} \sup_{||f_2||_\infty,\ldots,||f_k||_\infty\leq 1} \sup_{|c_{w,r}|\leq 1} \Big\| \mathop{\mathbb{E}}_{0\leq w\leq D_r} c_{w,r}\, T^{\lfloor P_{1,r,v_r}(w)\rfloor}\tilde{f}_r \cdot \ldots \cdot T^{\lfloor P_{k,r,v_r}(w)\rfloor} f_k \Big\|_{L^2(\mu)}^{2^t} = 0. \quad (3.50)$$

We also choose functions $h_{1,R},\ldots,h_{\ell,R}\in L^\infty(\mu)$ so that the corresponding average is $1/R$ close to the supremum of the $h_1,\ldots,h_\ell$. Namely, we want to prove (3.50) where f_r is now the function

$$f_1 \cdot T^{\lfloor\theta_1(r)\rfloor}h_{1,R}\cdot\ldots\cdot T^{\lfloor\theta_\ell(r)\rfloor}h_{\ell,R}.$$

3.7.6 Step 6: Applying the polynomial bounds

Now, we apply Proposition 3.4.5 for the inner average in the above relation. We have established that its hypotheses are satisfied. The degree and the type of the polynomial family all depend on the initial functions $a_1,\ldots,a_k$. Therefore, all asymptotic bounds are assumed to depend only on $a_1,\ldots,a_k$ and we omit the indices.

Let us denote the leading vector of the family $\mathcal{P}_r$ by $(u_{1,r},\ldots,u_{k,r})$ and recall again here that each $u_{i,r}$ satisfies one of the following:
a) there exists a sub-linear function $\phi_i(r)\prec u(r)$ that dominates some fractional power, such that $u_{i,r}=\phi_i(r)+o_r(1)$,
b) they have the form $c\lfloor u(r)\rfloor^\theta$, where $c\in\mathbb{R}$ and θ is a positive integer or
c) they converge to a non-zero real number.

Fix a positive integer M. There exist integers s,t, a finite set Y of integers and polynomials $p_{\underline{\varepsilon},i}$ (all depending only on the original functions $a_1,\ldots,a_k$), where $\underline{\varepsilon}\in[[s]]$ and $1\leq i\leq k$ such that

$$\sup_{||f_2||_\infty\leq 1,\ldots,||f_k||_\infty\leq 1} \sup_{|c_{w,r}|\leq 1} \Big\| \mathop{\mathbb{E}}_{0\leq w\leq D_r} c_{w,r}\, T^{\lfloor P_{1,r,v_r}(w)\rfloor}\tilde{f}_r \cdot \ldots \cdot T^{\lfloor P_{k,r,v_r}(w)\rfloor} f_k \Big\|_{L^2(\mu)}^{2^t} \ll \frac{1}{M} + \sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} T^{\lfloor A_{\underline{\varepsilon},r}(\mathbf{m})\rfloor + h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}\tilde{f}_r)\, d\mu \Big| + o_r(1), \quad (3.51)$$

where

$$A_{\underline{\varepsilon},r}(\mathbf{m}) = \sum_{1\leq j\leq k} p_{\underline{\varepsilon},j}(\mathbf{m})u_{j,r}. \quad (3.52)$$

The polynomials $A_{\underline{\varepsilon}}$ are essentially distinct for any value of the $u_{j,r}$ and satisfy

$$A_{\underline{\varepsilon},r}(\mathbf{m}) + A_{\underline{\varepsilon}^c,r}(\mathbf{m}) = A_{\underline{1},r}(\mathbf{m}).$$

In addition, for an $\underline{\varepsilon}\in[[s]]$, we have that the non-zero polynomials among the $p_{\underline{\varepsilon},j}$ are linearly independent.

Applying the bounds of (3.51) to (3.50), we deduce that our original average is bounded by the quantity

$$\frac{1}{M} + \sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \mathop{\mathbb{E}}_{1\leq r\leq R} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} T^{\lfloor A_{\underline{\varepsilon},r}(\mathbf{m})\rfloor + h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}\tilde{f}_r)\, d\mu \Big| + o_R(1) =$$
$$\frac{1}{M} + \sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \mathop{\mathbb{E}}_{1\leq r\leq R} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\leq i\leq \ell} T^{\lfloor A_{\underline{\varepsilon},r}(\mathbf{m})\rfloor + \lfloor\theta_i(r)\rfloor + h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}h_{i,R})\, d\mu \Big| + o_R(1), \quad (3.53)$$

where we set $\theta_0(r)\equiv 0$ and $h_{0,R}\equiv f_1$ for convenience in notation. We may assume without loss of generality that $0\equiv\theta_0(r)\ll\theta_1(r)\ll\ldots\ll\theta_\ell(r)$. Then, we compose with $T^{-\lfloor\theta_\ell(r)\rfloor}$ inside the above integral and combine the integer parts to obtain that the aforementioned integral is equal to

$$\int \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\leq i\leq\ell} T^{\lfloor A_{\underline{\varepsilon},r}(\mathbf{m})\rfloor + \lfloor\theta_i(r)-\theta_\ell(r)\rfloor + h_{i,r} + h_{\underline{\varepsilon}}}(\mathcal{C}^{|\underline{\varepsilon}|}h_{i,R})\, d\mu,$$

where $h_{i,r} \in \{0, \pm 1\}$. Putting this in (3.53), we want to bound

$$\frac{1}{M} + \sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \mathop{\mathbb{E}}_{1\le r\le R} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\le i\le \ell} T^{\lfloor A_{\underline{\varepsilon},r}(\mathbf{m})\rfloor + \lfloor \theta_i(r)-\theta_\ell(r)\rfloor + h_{i,r} + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} h_{i,R}) \ d\mu \Big| + o_R(1).$$

Using the argument present in Lemma 3.2.1, we deduce that the last quantity is smaller than a constant multiple of

$$\frac{1}{M} + \sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \sup_{\|h_1\|_\infty,\dots,\|h_\ell\|_\infty\le 1} \mathop{\mathbb{E}}_{1\le r\le R} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\le i\le \ell} T^{\lfloor A_{\underline{\varepsilon},r}(\mathbf{m})\rfloor + \lfloor \theta_i(r)-\theta_\ell(r)\rfloor + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} h_i) \ d\mu \Big| + o_R(1).$$

We choose again sequences of functions in place of the $h_1, \dots, h_\ell$, so that the corresponding quantity is $1/R$ close to the supremum and we denote them again $h_{1,R}, \dots, h_{\ell,R}$ for convenience. Note that this final quantity is essentially has the same form as the one in (3.53), but the function θ_0 corresponding to f_1 now has maximal growth rate among the θ_i. Therefore, our original problem reduces to finding a bound for

$$\frac{1}{M} + \sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \mathop{\mathbb{E}}_{1\le r\le R} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\le i\le \ell} T^{\lfloor A_{\underline{\varepsilon},r}(\mathbf{m})\rfloor + \lfloor \theta_i(r)\rfloor + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} h_{i,R}) \quad d\mu \Big| + o_R(1) \tag{3.54}$$

under the assumption that $\theta_0(t) \gg \theta_i(t) \succ \log t$ for every $1 \le i \le l-1$, $\theta_\ell \equiv 0$ and $\theta_i(t) - \theta_j(t) \succ \log t$ for all $i \ne j$.

We write

$$B_{\mathbf{m},\mathbf{h}}(r) := \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\le i\le \ell} T^{\lfloor A_{\underline{\varepsilon},r}(\mathbf{m})\rfloor + \lfloor \theta_i(r)\rfloor + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} h_{i,R}) \ d\mu \Big|.$$

Taking the limit as $R \to +\infty$, our goal is to show that the quantity

$$\frac{1}{M} + \sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \big(\limsup_{R\to+\infty} \mathop{\mathbb{E}}_{1\le r\le R} B_{\mathbf{m},\mathbf{h}}(r) \big).$$

goes to 0, as M goes to infinity.

3.7.7 Step 7: Another change of variables trick

Before we proceed with the final details of the proof, we will make a final trick to reduce our problem to a statement, where the results of Section 3.6 can be applied. We will use Lemma 3.2.3 to achieve this.

Our main objective is the following: since the sequences $u_{j,r}$ of the leading vector can have the form $c\lfloor u(r)\rfloor^k$, which are tough to handle, we want to use the above lemma to replace these terms with the terms cr^k, which are just polynomials. In order to facilitate this, we need to write the entire integral $B_{\mathbf{m},\mathbf{h}}(r)$ as a function of $\lfloor u(r)\rfloor$. Note that $u(r)$ satisfies the growth condition in the statement of Lemma 3.2.3. We consider three cases:

i) If the sequence $u_{j,r}$ has the form $c\lfloor u(r)\rfloor^q$, for $c \in \mathbb{R}$ and $q \in \mathbb{N}^*$, then it is already written as a function of $\lfloor u(r)\rfloor$.

ii) If the sequence $u_{j,r}$ converges to a non-zero real number a_j, then, we have $u_{j,r} - a_j = o_r(1)$ and the constant function a_j is already written as a function of $\lfloor u(r)\rfloor$.

iii) Finally, assume the sequence $u_{j,r}$ satisfies the remaining possible condition, namely that there exists a function $\phi_j \in \mathcal{H}$ satisfying the growth condition

$$t^\delta \prec \phi_j(t) \prec u(t)$$

for some $\delta > 0$ and such that

$$u_{j,r} = \phi_j(r) + o_r(1).$$

Let us assume that $\phi_j(t)$ is eventually positive (in the other case, we work with the number $-u_{j,r}$). We write $\phi_j(t) = \Phi_j(u(t))$, where $\Phi_j = \phi_j \circ u^{-1}$, which is well defined and thus a function in $\mathcal{H}$ [13]. We also have that $\Phi_j(t) \prec t$ (this follows easily from the fact that $\phi_j(t) \prec u(t)$) and we can easily see that $\Phi_j(t)$ also dominates some fractional power. In addition, we have

$$|\Phi_j(u(t)) - \Phi_j(\lfloor u(t) \rfloor)| \leq \sup_{x\in\mathbb{R}, \lfloor u(t)\rfloor \leq x \leq u(t)} |\Phi_j'(t)| = o_t(1),$$

since $\Phi_j'(t) \ll \Phi_j(t)/t \prec 1$.

In all three cases above, we have the following: there exists a function $w_j \in \mathcal{H}$, such that

$$|u_{j,r} - w_j(\lfloor u(r) \rfloor)| = o_r(1) \tag{3.55}$$

and the function w_j is either a monomial, or a constant function or a sub-linear (but not a sub-fractional) function. We write

$$\tilde{A}_{\underline{\varepsilon},r}(\mathbf{m}) = \sum_{1\leq j\leq k} p_{\underline{\varepsilon},j}(\mathbf{m}) w_j(\lfloor u(r) \rfloor) \tag{3.56}$$

and observe that $|A_{\underline{\varepsilon},r}(\mathbf{m}) - \tilde{A}_{\underline{\varepsilon},r}(\mathbf{m})| = o_r(1)$, for any fixed value of $\mathbf{m}$. Therefore, for r large enough, we have

$$\lfloor A_{\underline{\varepsilon},r}(\mathbf{m}) \rfloor = \left\lfloor \tilde{A}_{\underline{\varepsilon},r}(\mathbf{m}) \right\rfloor + h'_{r,\underline{\varepsilon},\mathbf{m}}, \tag{3.57}$$

where $h'_{r,\underline{\varepsilon},\mathbf{m}} \in \{0, \pm 1\}$.

We do the same for the function θ_i. Indeed, we can use the same arguments as above to deduce that $|\theta_i(t) - \psi_i(\lfloor u(t) \rfloor)| = o_t(1)$, where $\psi_i(t) \in \mathcal{H}$ is the function $\theta_i \circ u^{-1}$ In addition, since u dominates some fractional power, we have that u^{-1} has polynomial growth and, therefore, we easily get $t^{\varepsilon} \succ \psi_i(t) \succ \log t$ for all $\varepsilon > 0$, that is ψ_i is a (sub-fractional) function. Finally, for r large enough, we can write

$$\lfloor \theta_i(r) \rfloor = \lfloor \psi_i(u(r)) \rfloor + h''_{i,r}, \tag{3.58}$$

where $h''_{i,r} \in \{0, \pm 1\}$.

In view of the above, we have

$$\begin{aligned}
&\sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \mathop{\mathbb{E}}_{1\leq r\leq R} B_{\mathbf{m},\mathbf{h}}(r) = \\
&\sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \mathop{\mathbb{E}}_{1\leq r\leq R} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\leq i\leq \ell} T^{\lfloor \tilde{A}_{\underline{\varepsilon},r}(\mathbf{m}) \rfloor + h'_{r,\underline{\varepsilon},\mathbf{m}} + \lfloor \psi_i(\lfloor u(r)\rfloor)\rfloor + h''_{i,r} + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} h_{i,R}) \, d\mu \Big| + o_R(1) \leq \\
&\quad \sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \Big(\mathop{\mathbb{E}}_{1\leq r\leq R} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\leq i\leq \ell} T^{\lfloor \tilde{A}_{\underline{\varepsilon},r}(\mathbf{m}) \rfloor + h'_{r,\underline{\varepsilon},\mathbf{m}} + \lfloor \psi_i(\lfloor u(r)\rfloor)\rfloor + h''_{i,r} + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} h_{i,R}) \, d\mu \Big|^2 \Big)^{1/2} \\
&\qquad + o_R(1),
\end{aligned}$$

where we applied the Cauchy-Schwarz inequality (the $o_R(1)$ term on the second line exists to account for small values of r for which (3.57),(3.58) may not hold with error terms in the set $\{0, \pm 1\}$). Thus, we want to bound

$$\frac{1}{M} + \sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \Big(\mathop{\mathbb{E}}_{1\leq r\leq R} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\leq i\leq \ell} T^{\lfloor \tilde{A}_{\underline{\varepsilon},r}(\mathbf{m}) \rfloor + h'_{r,\underline{\varepsilon},\mathbf{m}} + \lfloor \psi_i(\lfloor u(r)\rfloor)\rfloor + h''_{i,r} + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} h_{i,R}) \, d\mu \Big|^2 \Big)^{1/2} + o_R(1), \tag{3.59}$$

where $h_{0,R} = f_1$.

[13] Note that $u(t)$ is a positive function by its definition and therefore, goes to $+\infty$. Consequently, u^{-1} also goes to $+\infty$.

Claim 6. *Proposition 3.1.1 holds in the case when all the functions w_j (defined in (3.55)) are constant and $\ell = 0$.*

Proof of the claim. This means that the polynomials $\tilde{A}_{\underline{\varepsilon},r}(\mathbf{m})$ are actually independent of r and we write them as $\tilde{A}_{\underline{\varepsilon}}(\mathbf{m})$. In addition, there are no functions ψ_i in the iterates of the above quantity. Finally, the error terms $h''_{i,r}$ do not exist in this case. Our problem reduces to finding a bound for

$$\frac{1}{M} + \sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \Big(\mathop{\mathbb{E}}_{1\leq r\leq R} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} T^{\lfloor \tilde{A}_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h'_{r,\underline{\varepsilon},\mathbf{m}} + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} f_1)\, d\mu \Big|^2 \Big)^{1/2} + o_R(1), \tag{3.60}$$

where $h'_{r,\underline{\varepsilon},\mathbf{m}} \in \{0, \pm 1\}$. Note that

$$\mathop{\mathbb{E}}_{1\leq r\leq R} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} T^{\lfloor \tilde{A}_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h'_{r,\underline{\varepsilon},\mathbf{m}} + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} f_1)\, d\mu \Big|^2 \leq$$

$$\sum_{h'_{\underline{\varepsilon}}\in\{0,\pm1\},\underline{\varepsilon}\in[[s]]} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} T^{\lfloor \tilde{A}_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h'_{\underline{\varepsilon}} + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} f_1)\, d\mu \Big|^2,$$

which implies that the quantity in (3.60) is smaller than $O(1)$ times

$$\frac{1}{M} + \sum_{\mathbf{h}\in \tilde{Y}^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} T^{\lfloor \tilde{A}_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} f_1)\, d\mu \Big| + o_R(1)$$

for some new, larger finite set $\tilde{Y}$. The statement follows if we prove that

$$\lim_{M\to+\infty} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} T^{\lfloor \tilde{A}_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} f_1)\, d\mu \Big| = 0$$

for any $h_{\underline{\varepsilon}} \in \mathbb{Z}$. Note that the polynomials $\tilde{A}_{\underline{\varepsilon}}(\mathbf{m})$ are essentially distinct as the statement of Proposition 3.4.5 guarantees. Squaring and applying the Cauchy-Schwarz inequality, we want to prove that

$$\lim_{M\to+\infty} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} T^{\lfloor \tilde{A}_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} f_1)\, d\mu \Big|^2 = 0,$$

which can be rewritten as

$$\lim_{M\to+\infty} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \int \prod_{\underline{\varepsilon}\in[[s]]} S^{\lfloor \tilde{A}_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} F_1)\, d(\mu\times\mu) = 0,$$

where $S = T \times T$ and $F_1 = \overline{f_1} \otimes f_1$. This is an average where the iterates are real polynomials and using [17, Lemma 4.3], we can prove that this last relation holds, provided that $\|S^{h_{\underline{1}}} F_1\|_{\tilde{s},T\times T} = 0$, for some positive integer $\tilde{s}$ that depends only on the polynomials $A_{\underline{\varepsilon}}$ (which depend on the original Hardy field functions $a_1, ..., a_k$). However, since $\|F_1\|_{\tilde{s},T\times T} \leq \|f_1\|^2_{\tilde{s}+1,T}$, we get that the statement holds if the function f_1 satisfies $\|f_1\|_{\tilde{s}+1,T} = 0$. This completes the proof of our claim. □

From now on, we assume that either at least one of the functions w_j is non-constant, or that $\ell \geq 1$ and we want to bound the quantity in (3.59). Writing $H_{i,R} = \overline{h_{i,R}} \otimes h_{i,R}$ and $S = T \times T$, we observe that

$$\mathop{\mathbb{E}}_{1\leq r\leq R} \Big| \int \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\leq i\leq \ell} T^{\lfloor \tilde{A}_{\underline{\varepsilon},r}(\mathbf{m})\rfloor + h'_{r,\underline{\varepsilon},\mathbf{m}} + \lfloor \psi_i(\lfloor u(r)\rfloor)\rfloor + h''_{i,r} + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} h_{i,R})\, d\mu \Big|^2 =$$

$$\mathop{\mathbb{E}}_{1\leq r\leq R} \int \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\leq i\leq \ell} S^{\lfloor \tilde{A}_{\underline{\varepsilon},r}(\mathbf{m})\rfloor + h'_{r,\underline{\varepsilon},\mathbf{m}} + \lfloor \psi_i(\lfloor u(r)\rfloor)\rfloor + h''_{i,r} + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} H_{i,R})\, d(\mu\times\mu) \leq$$

$$\Big\| \mathop{\mathbb{E}}_{1\leq r\leq R} \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\leq i\leq \ell} S^{\lfloor \tilde{A}_{\underline{\varepsilon},r}(\mathbf{m})\rfloor + h'_{r,\underline{\varepsilon},\mathbf{m}} + \lfloor \psi_i(\lfloor u(r)\rfloor)\rfloor + h''_{i,r} + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} H_{i,R}) \Big\|_{L^2(\mu\times\mu)}$$

due to the Cauchy-Schwarz inequality. Invoking[14] Lemma 3.2.1, we have

$$\Big\| \mathop{\mathbb{E}}_{1\le r\le R} \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\le i\le \ell} S^{\lfloor \tilde{A}_{\underline{\varepsilon},r}(\mathbf{m})\rfloor + h'_{r,\underline{\varepsilon},\mathbf{m}} + \lfloor \psi_i(\lfloor u(r)\rfloor)\rfloor + h''_{i,r} + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} H_{i,R}) \Big\|_{L^2(\mu\times\mu)} \ll_{s,\ell}$$

$$\sup_{|c_{r,\mathbf{m},\mathbf{h}}|\le 1} \sup_{\|H_i\|_\infty \le 1} \Big\| \mathop{\mathbb{E}}_{1\le r\le R} c_{r,\mathbf{m},\mathbf{h}} \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\le i\le \ell} S^{\lfloor \tilde{A}_{\underline{\varepsilon},r}(\mathbf{m})\rfloor + \lfloor \psi_i(\lfloor u(r)\rfloor)\rfloor + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} H_i) \Big\|_{L^2(\mu\times\mu)},$$

where $H_0 = \overline{f_1} \otimes f_1$ and $\mathbf{h} = (h_{\underline{\varepsilon}}, \underline{\varepsilon} \in [[s]])$. Note that since both s, ℓ depend on the original Hardy field functions $a_1, ..., a_k$, the implicit constant in the last bound depends only on $a_1, ..., a_k$ (which we omit from the subscripts).

Putting everything together, we get that

$$\frac{1}{M} + \sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \mathop{\mathbb{E}}_{1\le r\le R} B_{\mathbf{m},\mathbf{h}}(r) + o_R(1)$$

$$\ll \frac{1}{M} + \sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \sup_{|c_{r,\mathbf{m},\mathbf{h}}|\le 1} \sup_{\substack{\|H_i\|_\infty \le 1\\ 1\le i\le \ell}}$$

$$\Big\| \mathop{\mathbb{E}}_{1\le r\le R} c_{r,\mathbf{m},\mathbf{h}} \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\le i\le \ell} S^{\lfloor \tilde{A}_{\underline{\varepsilon},r}(\mathbf{m})\rfloor + \lfloor \psi_i(\lfloor u(r)\rfloor)\rfloor + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} H_i) \Big\|^{1/2}_{L^2(\mu\times\mu)} + o_R(1).$$

Now, we choose functions $H_{1,R}, ..., H_{\ell,R}$ so that the above average (over R) is $1/R$ close to the supremum. Then, we take the limit as $R \to +\infty$ and apply Lemma 3.2.3 to deduce that the limsup of this last quantity is bounded by $O_u(1)$ times (which is, of course, $O_{a_1,...,a_k}(1)$)

$$\sum_{\mathbf{h}\in Y^{[[s]]}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \limsup_{R\to+\infty} \sup_{|c_{r,\mathbf{m},\mathbf{h}}|\le 1}$$

$$\Big\| \mathop{\mathbb{E}}_{1\le r\le R} c_{r,\mathbf{m},\mathbf{h}} \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\le i\le \ell} S^{\lfloor \widehat{A}_{\underline{\varepsilon},r}(\mathbf{m})\rfloor + \lfloor \psi_i(r)\rfloor + h_{\underline{\varepsilon}}} (\mathcal{C}^{|\underline{\varepsilon}|} H_{i,R}) \Big\|^{1/2}_{L^2(\mu\times\mu)},$$

where we define (recall (3.56))

$$\widehat{A}_{\underline{\varepsilon},r}(\mathbf{m}) := \sum_{1\le j\le k} p_{\underline{\varepsilon},j}(\mathbf{m}) w_j(r).$$

and $H_{0,R} = \overline{f_1} \otimes f_1$. Finally, we can combine the integer parts in the iterates of the above quantity (using again Lemma 3.2.1 to remove the error terms). In conclusion, our original average is bounded by $O(1)$ times

$$\frac{1}{M} + \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \Big(\limsup_{R\to+\infty} \sup_{|c_{r,\mathbf{m}}|\le 1} \Big\| \mathop{\mathbb{E}}_{1\le r\le R} c_{r,\mathbf{m}} \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\le i\le \ell} S^{\lfloor \widehat{A}_{\underline{\varepsilon},r}(\mathbf{m}) + \psi_i(r)\rfloor} (\mathcal{C}^{|\underline{\varepsilon}|} H_{i,R}) \Big\|^{1/2}_{L^2(\mu\times\mu)} \Big) \le$$

$$\frac{1}{M} + \Big(\mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \limsup_{R\to+\infty} \sup_{|c_{r,\mathbf{m}}|\le 1} \Big\| \mathop{\mathbb{E}}_{1\le r\le R} c_{r,\mathbf{m}} \prod_{\underline{\varepsilon}\in[[s]]} \prod_{0\le i\le \ell} S^{\lfloor \widehat{A}_{\underline{\varepsilon},r}(\mathbf{m}) + \psi_i(r)\rfloor} (\mathcal{C}^{|\underline{\varepsilon}|} H_{i,R}) \Big\|_{L^2(\mu\times\mu)} \Big)^{1/2} \tag{3.61}$$

by the Cauchy-Schwarz inequality. Note that all implied asymptotic constants above did not depend on either M or R.

[14] Note that all the error terms depending on r in the iterates take values on finite sets.

3.7.8 Finishing the proof

We describe the final step here. Our main observation is that $\widehat{A}_{\underline{\varepsilon},r}(\mathbf{m})+\psi_i(r)$, when viewed as a function of r, is a sum of sub-linear functions that dominate the function $\log r$ and monomials (possibly of degree 0). Our goal is to use the bounds in Proposition 3.6.1 to deduce our result. However, it is not immediately obvious that in our case a linear combination of functions of the above form dominates the logarithmic function $\log r$ (the statement in general is false and a counterexample is given by the pair $(\log^2 r+\log r, \log^2 r)$). We shall establish that this is true for all $\mathbf{m}\in\mathbb{Z}^t$ outside of a negligible set. We recall here that for every large enough r (large enough for $w_j(r)$ to be non-zero), the $\widehat{A}_{\underline{\varepsilon},r}(\mathbf{m})$ are pairwise essentially distinct polynomials in the variable $\mathbf{m}$ and in addition satisfy

$$\widehat{A}_{\underline{\varepsilon},r}(\mathbf{m})+\widehat{A}_{\underline{\varepsilon}^c,r}(\mathbf{m})=\widehat{A}_{\underline{1},r}(\mathbf{m}).$$

We will use the following lemma:

Lemma 3.7.3. *Let $p\in\mathbb{R}^t(\mathbf{x})$ be a non-zero real polynomial of degree d. Then, the set of integer solutions of the equation*

$$p(\mathbf{m})=0$$

in $[-M,M]^t$ has $O_d(M^{t-1})$ elements.

Proof. For $t=1$ it is obvious, since the polynomial has at most d roots. Assume we have proven the result for $t-1$. We can write $p(\mathbf{m})$ in the form

$$p(m_1,...,m_t)=a_{d'}(m_1,...,m_{t-1})m_t^{d'}+\cdots+a_1(m_1,...,m_{t-1})m_t+a_0(m_1,....,m_{t-1})$$

for some $d'\leq d$. At least one of the polynomials $a_i(m_1,...,m_{t-1})$ with $1\leq i\leq d'$ is not identically zero and thus has at most $O_{d,t}(M^{t-2})$ zeroes in $[-M,M]^{t-1}$. If $(x_1,...,x_{t-1})$ is not one of these zeroes, then $p(x_1,...,x_{t-1},m_t)$ is non-trivial as a polynomial in the variable m_t. Therefore, it is satisfied by no more than d values of m_t. Summing over all tuples $(m_1,...,m_{t-1})\in[-M,M]^{t-1}$, we get the result. □

Corollary 3.7.4. *Let $a_1\ll...\ll a_k$ be functions in $\mathcal{H}$ and let $p_1(\mathbf{m}),...,p_k(\mathbf{m})\in\mathbb{R}^t(\mathbf{x})$ be non-zero linearly independent polynomials. Then, for all $\mathbf{m}\in\mathbb{Z}^t$ outside a set of density 0, we have that*

$$p_1(\mathbf{m})a_1+\cdots+p_k(\mathbf{m})a_k\sim a_k. \tag{3.62}$$

Proof. Let $a_{k_0},...,a_k$ be the functions among the a_i that have the same growth rate as a_k. Then, for $k_0\leq j\leq k$, we can write $a_j(t)=c_ja_k(t)+b_j(t)$, where $c_j\in\mathbb{R}^*$ and $b_j(t)\prec a_k(t)$. Then, the function in (3.62) has the same growth rate as the function

$$\big(c_{k_0}p_{k_0}(\mathbf{m})+\cdots+c_kp_k(\mathbf{m})\big)a_k(t)$$

unless of course $c_{k_0}p_{k_0}(\mathbf{m})+\cdots+c_kp_k(\mathbf{m})=0$. However, the linear independence hypothesis implies that this polynomial is non-zero, and thus the set of of $\mathbf{m}\in\mathbb{Z}^t$ for which this last relation holds has density 0 in $\mathbb{Z}^t$ by Lemma 3.7.3. The conclusion follows. □

We use this corollary to prove the following:

Claim 7. *For all $\mathbf{m}\in\mathbb{Z}^t$ outside a set Λ of density 0, we have that the functions (in the variable r)*

$$\widehat{A}_{\underline{\varepsilon},r}(\mathbf{m})+\psi_i(r)=\sum_{1\leq j\leq k}p_{\underline{\varepsilon},j}(\mathbf{m})w_j(r)+\psi_i(r)$$

are a sum of a sub-linear function and a real polynomial. In addition, we have that they either dominate the function $\log r$, or they are a constant function.

Proof of the claim. We use Corollary 3.7.4 to find a set $\Lambda \subset \mathbb{Z}^t$ of density zero, so that for $\mathbf{m} \notin \Lambda$, we have that for any $\underline{\varepsilon} \in [[s]]$ and any subcollection J of the indices $j \in \{1, 2, ..., k\}$, we have that

$$\sum_{j \in J} p_{\underline{\varepsilon},j}(\mathbf{m}) w_j(r) \sim w_{\max(J)}(r),$$

where $w_{\max(J)}$ denotes a function in the collection $\{w_j, j \in J\}$ that has maximal growth rate. We show that this set Λ is sufficient for the statement of the claim to hold.

We split the w_j into two sets: the set S_1 consists of those functions that are monomials, while S_2 contains the rest (namely the sub-linear functions). Reordering, if necessary, we may assume that $S_1 = \{w_1, ..., w_{k_0}\}$ while $S_2 = \{w_{k_0+1}, ..., w_k\}$. We write

$$\widehat{A}_{\underline{\varepsilon},r}(\mathbf{m}) = \sum_{j=1}^{k_0} p_{\underline{\varepsilon},j}(\mathbf{m}) w_j(r) + \sum_{j=k_0+1}^{k} p_{\underline{\varepsilon},j}(\mathbf{m}) w_j(r). \tag{3.63}$$

For a fixed $\mathbf{m} \notin \Lambda$, the first summand is a polynomial in the variable r (possibly constant), while the second is a sub-linear function of r. Since the sub-linear functions w_j with $k_0 + 1 \leq j \leq k$ dominate some fractional power, we deduce that $\widehat{A}_{\underline{\varepsilon},r}(m)$ is either a constant function [15], or the sum of a polynomial and a sub-linear function that dominates some fractional power, since

$$\sum_{j=k_0+1}^{k} p_{\underline{\varepsilon},j}(\mathbf{m}) w_j(r) \sim w_{\max\{k_0+1,...,k\}}(r)$$

where $w_{\max(S)}$ for $S \subseteq \{1, ..., k\}$ is defined above and this is a sub-linear (but not sub-fractional) function.

In addition, if $\psi_i(t) \neq \psi_\ell(t)$ (recall that $\psi_\ell(t) \equiv 0$), we can use the same argument to show that

$$\widehat{A}_{\underline{\varepsilon},r}(\mathbf{m}) + \psi_i(r)$$

is a sum of a sub-linear function that dominates $\log r$ and a polynomial (we use the fact that ψ_i and w_j (for any j) have distinct growth rates, since the ψ_i is a sub-fractional function.) □

Let $\Lambda \subset \mathbb{Z}^t$ be the zero density set given by the above claim. Now, we isolate the iterate $S^{\lfloor \widehat{A}_{\underline{1},r}(\mathbf{m}) + \psi_0(t) \rfloor}(\mathcal{C}^{|\underline{1}|} H_0)$ in (3.61) and we also assume that $\mathbf{m} \notin \Lambda$. The above proof implies that the Hardy field function involved in this iterate is a sum of a sub-linear function (that dominates the logarithm) and a polynomial. In order to apply the results of Section 3.6, we have to show that the differences of this function with the rest of the functions in the iterates satisfies the same condition. That is, for every $(\underline{\varepsilon}, i) \neq (\underline{1}, 0)$, we have to show that the function

$$\big(\widehat{A}_{\underline{1},r}(\mathbf{m}) + \psi_0(r)\big) - \big(A_{\underline{\varepsilon},r}(\mathbf{m}) + \psi_i(r)\big)$$

is a sub-linear function plus a polynomial, or is bounded. Rewrite the above as

$$\widehat{A}_{\underline{\varepsilon}^c,r}(\mathbf{m}) + (\psi_0(r) - \psi_i(r)).$$

If $i \neq 0$, then we use the fact that $\psi_0 - \psi_i \succ \log t$ and the argument of the previous proof to establish that

$$\big(\widehat{A}_{\underline{1},r}(\mathbf{m}) + \psi_0(r)\big) - \big(\widehat{A}_{\underline{\varepsilon},r}(\mathbf{m}) + \psi_i(r)\big) \succ \log r$$

for all $\mathbf{m}$ outside a zero density set (which we attach to the set Λ) and that this function is the sum of a polynomial and a sub-linear function.

If $i = 0$, then the above difference is equal to $\widehat{A}_{\underline{\varepsilon}^c,r}(\mathbf{m})$ which is either the sum of a polynomial and a sub-linear function (that dominates $\log r$), or a constant function of r. We use this characterization

[15] This is the case when $p_{\underline{\varepsilon},j}(\mathbf{m}) \equiv 0$ for $j \geq k_0 + 1$ and the monomials w_j are constant polynomials in the variable r.

to split $[[s]]$ into two subsets: A_2 contains those $\underline{\varepsilon} \in [[s]]$, for which $\widehat{A}_{\underline{\varepsilon}^c,r}(\mathbf{m})$ satisfies the first condition, while the set A_1 contains the rest. Note that if $\underline{\varepsilon} \in A_1$, then the difference

$$\big(\widehat{A}_{\underline{1},r}(\mathbf{m}) + \psi_0(r)\big) - \big(\widehat{A}_{\underline{\varepsilon},r}(\mathbf{m}) + \psi_0(r)\big)$$

is a (non-constant) polynomial in the variable $\mathbf{m}$ and we denote it by $c_{\underline{\varepsilon}}(\mathbf{m})$. Thus, we can write

$$\big(\widehat{A}_{\underline{\varepsilon},r}(\mathbf{m}) + \psi_0(r)\big) = \big(\widehat{A}_{\underline{1},r}(\mathbf{m}) + \psi_0(r)\big) - c_{\underline{\varepsilon}}(\mathbf{m}).$$

Note that the polynomials $c_{\underline{\varepsilon}}(\mathbf{m})$ are essentially distinct, since the $\widehat{A}_{\underline{\varepsilon},r}$ are essentially distinct.

In view of the above, we rewrite the quantity in (3.61) as

$$\frac{1}{M} + \Big(\mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \limsup_{R\to+\infty} \sup_{|c_{r,\mathbf{m}}|\leq 1} \Big\| \mathop{\mathbb{E}}_{1\leq r\leq R} c_{r,\mathbf{m}} \prod_{\underline{\varepsilon}\in A_1} S^{\lfloor \widehat{A}_{\underline{1},r}(\mathbf{m})+\psi_0(r)-c_{\underline{\varepsilon}}(\mathbf{m})\rfloor}(\mathcal{C}^{|\underline{\varepsilon}|}H_0) \prod_{\underline{\varepsilon}\in A_2} S^{\lfloor \widehat{A}_{\underline{\varepsilon},r}(\mathbf{m})+\psi_0(r)\rfloor}(\mathcal{C}^{|\underline{\varepsilon}|}H_0) \prod_{1\leq i\leq \ell}\prod_{\underline{\varepsilon}\in[[s]]} S^{\lfloor \widehat{A}_{\underline{\varepsilon},r}(\mathbf{m})+\psi_i(r)\rfloor}(\mathcal{C}^{|\underline{\varepsilon}|}H_{i,R}) \Big\|_{L^2(\mu\times\mu)}\Big)^{1/2}. \tag{3.64}$$

Note that

$$\Big\lfloor \widehat{A}_{\underline{1},r}(\mathbf{m}) + \psi_0(r) - c_{\underline{\varepsilon}}(\mathbf{m})\Big\rfloor = \Big\lfloor \widehat{A}_{\underline{1},r}(\mathbf{m}) + \psi_0(r)\Big\rfloor + \big\lfloor -c_{\underline{\varepsilon}}(\mathbf{m})\big\rfloor + h_{\underline{\varepsilon},r,\mathbf{m}},$$

where $h_{\underline{\varepsilon},r,\mathbf{m}} \in \{0,\pm 1\}$. Thus, we rewrite (3.64) as

$$\frac{1}{M} + \Big(\mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \limsup_{R\to+\infty} \sup_{|c_{r,\mathbf{m}}|\leq 1} \Big\| \mathop{\mathbb{E}}_{1\leq r\leq R} c_{r,\mathbf{m}}\, S^{\lfloor \widehat{A}_{\underline{1},r}(\mathbf{m})+\psi_0(r)\rfloor}\Big(\prod_{\underline{\varepsilon}\in A_1} \mathcal{C}^{|\underline{\varepsilon}|} S^{\lfloor -c_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon},r,\mathbf{m}}} H_0\Big) \prod_{\underline{\varepsilon}\in A_2} S^{\lfloor \widehat{A}_{\underline{\varepsilon},r}(\mathbf{m})+\psi_0(r)\rfloor}(\mathcal{C}^{|\underline{\varepsilon}|}H_0) \prod_{1\leq i\leq \ell}\prod_{\underline{\varepsilon}\in[[s]]} S^{\lfloor \widehat{A}_{\underline{\varepsilon},r}(\mathbf{m})+\psi_i(r)\rfloor}(\mathcal{C}^{|\underline{\varepsilon}|}H_{i,R}) \Big\|_{L^2(\mu\times\mu)}\Big)^{1/2}. \tag{3.65}$$

Since $h_{\underline{\varepsilon},r,\mathbf{m}}$ take values in $\{0,\pm 1\}$, we can use the argument in Lemma 3.2.1 to deduce that

$$\Big\| \mathop{\mathbb{E}}_{1\leq r\leq R} c_{r,\mathbf{m}}\, S^{\lfloor \widehat{A}_{\underline{1},r}(\mathbf{m})+\psi_0(r)\rfloor}\Big(\prod_{\underline{\varepsilon}\in A_1} \mathcal{C}^{|\underline{\varepsilon}|} S^{\lfloor -c_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon},r,\mathbf{m}}} H_0\Big) \prod_{\underline{\varepsilon}\in A_2} S^{\lfloor \widehat{A}_{\underline{\varepsilon},r}(\mathbf{m})+\psi_0(r)\rfloor}(\mathcal{C}^{|\underline{\varepsilon}|}H_0) \prod_{1\leq i\leq \ell}\prod_{\underline{\varepsilon}\in[[s]]} S^{\lfloor \widehat{A}_{\underline{\varepsilon},r}(\mathbf{m})+\psi_i(r)\rfloor}(\mathcal{C}^{|\underline{\varepsilon}|}H_{i,R}) \Big\|_{L^2(\mu\times\mu)} \leq$$
$$\sum_{\substack{h_{\underline{\varepsilon},\mathbf{m}}\in\{0,\pm1\} \\ \underline{\varepsilon}\in A_1}} \sup_{|c'_{r,\mathbf{m}}|\leq 1} \Big\| \mathop{\mathbb{E}}_{1\leq r\leq R} c'_{r,\mathbf{m}}\, S^{\lfloor \widehat{A}_{\underline{1},r}(\mathbf{m})+\psi_0(r)\rfloor}\Big(\prod_{\underline{\varepsilon}\in A_1} \mathcal{C}^{|\underline{\varepsilon}|} S^{\lfloor -c_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon},\mathbf{m}}} H_0\Big) \prod_{\underline{\varepsilon}\in A_2} S^{\lfloor \widehat{A}_{\underline{\varepsilon},r}(\mathbf{m})+\psi_0(r)\rfloor}(\mathcal{C}^{|\underline{\varepsilon}|}H_0) \prod_{1\leq i\leq \ell}\prod_{\underline{\varepsilon}\in[[s]]} S^{\lfloor \widehat{A}_{\underline{\varepsilon},r}(\mathbf{m})+\psi_i(r)\rfloor}(\mathcal{C}^{|\underline{\varepsilon}|}H_{i,R}) \Big\|_{L^2(\mu\times\mu)}.$$

Thus, our problem reduces to showing that

$$\frac{1}{M} + \Big(\mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \limsup_{R\to+\infty} \sup_{|c_{r,\mathbf{m}}|\leq 1} \Big\| \mathop{\mathbb{E}}_{1\leq r\leq R} c_{r,\mathbf{m}}\, S^{\lfloor \widehat{A}_{\underline{1},r}(\mathbf{m})+\psi_0(r)\rfloor}\Big(\prod_{\underline{\varepsilon}\in A_1} \mathcal{C}^{|\underline{\varepsilon}|} S^{\lfloor -c_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon},\mathbf{m}}} H_0\Big) \prod_{\underline{\varepsilon}\in A_2} S^{\lfloor \widehat{A}_{\underline{\varepsilon},r}(\mathbf{m})+\psi_0(r)\rfloor}(\mathcal{C}^{|\underline{\varepsilon}|}H_0) \prod_{1\leq i\leq \ell}\prod_{\underline{\varepsilon}\in[[s]]} S^{\lfloor \widehat{A}_{\underline{\varepsilon},r}(\mathbf{m})+\psi_i(r)\rfloor}(\mathcal{C}^{|\underline{\varepsilon}|}H_{i,R}) \Big\|_{L^2(\mu\times\mu)}\Big)^{1/2} \tag{3.66}$$

goes to 0 as $M \to +\infty$ (that is, our error terms in the iterates do not depend on r now).

In order to be able to apply Proposition 3.6.1, we need to check that the degree, type and size (as defined in the beginning of Section 3.6) of the given collection of functions in the iterates is constant, as $\mathbf{m}$ ranges over $\mathbb{Z}^t$, so that we can use bounds that are uniform in the variable $\mathbf{m}$. Recall (3.63): the "polynomial component" of $\widehat{A}_{\underline{\varepsilon},r}(\mathbf{m}) + \psi_i(r)$ is

$$\sum_{j=1}^{k_0} p_{\underline{\varepsilon},j}(\mathbf{m}) w_j(r),$$

where the functions $w_j(r)$ are polynomials. The argument is as follows: for any two real polynomials $p_1(\mathbf{m})$ and $p_2(\mathbf{m})$ we must have that they are either equal for all $\mathbf{m}$, or the set of integer solutions of $p_1(\mathbf{m}) = p_2(\mathbf{m})$ has density zero. Comparing coefficients, it is straightforward to see that outside a set Λ' of density zero, the degree, type and size of the collection of functions in the iterates in (3.66) is independent of $\mathbf{m}$ for any $\mathbf{m} \notin \Lambda$ (and they all depend only on the initial Hardy field functions $a_1, ..., a_k$). In addition, the elements of the leading vector of this collection are polynomials in $\mathbf{m}$ (we are not concerned with their actual form here). Therefore, we write the leading vector as $(u_1(\mathbf{m}), ..., u_{s_0}(\mathbf{m}))$, where $s_0 \leq s$ is the size of the given collection of functions, which does not depend on $\mathbf{m}$ outside our "negligible" set. Furthermore, for $\mathbf{m}$ outside a set of density zero (which we attach to the set Λ'), we have that all the numbers $u_1(\mathbf{m}), ..., u_{s_0}(\mathbf{m})$ are non-zero, and thus we can now apply Proposition 3.6.1 for all $\mathbf{m}$ outside a negligible subset of $\mathbb{Z}^t$.

Write $\mathbf{h}_{\mathbf{m}} := (h_{\underline{\varepsilon},\mathbf{m}}, \underline{\varepsilon} \in A_1)$ and

$$F_{\mathbf{m},\mathbf{h}_{\mathbf{m}}} := \prod_{\underline{\varepsilon}\in A_1} \mathcal{C}^{|\underline{\varepsilon}|} S^{\lfloor -c_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon},\mathbf{m}}} H_0.$$

Now, for any $\mathbf{m} \notin \Lambda \cup \Lambda'$ we apply Proposition 3.6.1 (note we can have at most $2^s(\ell+1)$ different Hardy field functions in the iterates) to deduce that there exist positive integers t', s', a finite set $\widetilde{Y}$ and polynomials $p'_{\underline{\varepsilon},j}$, where $\underline{\varepsilon} \in [[s']]$ and $1 \leq j \leq s_0$ (which depend only on the original functions $a_1, ..., a_k$), such that

$$\limsup_{R\to+\infty} \sup_{|c_{r,\mathbf{m}}|\leq 1} \Big\| \mathop{\mathbb{E}}_{1\leq r\leq R} c_{r,\mathbf{m}}\, S^{\lfloor \widehat{A}_{1,r}(\mathbf{m})+\psi_0(r)\rfloor}\big(\prod_{\underline{\varepsilon}\in A_1} \mathcal{C}^{|\underline{\varepsilon}|} S^{\lfloor -c_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon},\mathbf{m}}} H_0\big)$$
$$\prod_{\underline{\varepsilon}\in A_2} S^{\lfloor \widehat{A}_{\underline{\varepsilon},r}(\mathbf{m})+\psi_0(r)\rfloor}(\mathcal{C}^{|\underline{\varepsilon}|} H_0) \prod_{1\leq i\leq \ell}\prod_{\underline{\varepsilon}\in[[s]]} S^{\lfloor \widehat{A}_{\underline{\varepsilon},r}(\mathbf{m})+\psi_i(r)\rfloor}(\mathcal{C}^{|\underline{\varepsilon}|} H_{i,R})\Big\|_{L^2(\mu\times\mu)}^{2^{t'}} \ll_{a_1,\dots,a_k}$$
$$\frac{1}{M} + \sum_{\mathbf{h}\in \widetilde{Y}^{[[s']]}} \mathop{\mathbb{E}}_{\mathbf{m}'\in[-M,M]^{t'}} \Big|\!\Big|\!\Big| \prod_{\underline{\varepsilon}'\in[[s']]} S^{\lfloor A_{\underline{\varepsilon}'}(\mathbf{m}',\mathbf{m})\rfloor + h_{\underline{\varepsilon}'}} F_{\mathbf{m},\mathbf{h}_{\mathbf{m}}} \Big|\!\Big|\!\Big|_{2^{s+1}(\ell+1),S}.$$

Here, we have defined

$$A_{\underline{\varepsilon}'}(\mathbf{m}',\mathbf{m}) = \sum_{j=1}^{s_0} p'_{\underline{\varepsilon},j}(\mathbf{m}') u_j(\mathbf{m}).$$

Therefore, since the set $\Lambda \cup \Lambda'$ has density zero, we use the Hölder inequality to get that the quantity in (3.66) is $\ll_{a_1,\dots,a_k}$

$$\mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \Big(\sum_{\mathbf{h}\in \widetilde{Y}^{[[s']]}} \mathop{\mathbb{E}}_{\mathbf{m}'\in[-M,M]^{t'}} \Big|\!\Big|\!\Big| \prod_{\underline{\varepsilon}'\in[[s']]} S^{\lfloor A_{\underline{\varepsilon}'}(\mathbf{m}',\mathbf{m})\rfloor + h_{\underline{\varepsilon}'}} F_{\mathbf{m},\mathbf{h}_{\mathbf{m}}} \Big|\!\Big|\!\Big|_{2^{s+1}(\ell+1),S}^{1/2^{t'}} \Big)^{1/2} + o_M(1).$$

Now, we take the limit as $M \to +\infty$ and use the power mean inequality to bound the lim sup of the above quantity by $O_{a_1,\dots,a_k}(1)$ times a power of

$$\limsup_{M\to+\infty} \sum_{\mathbf{h}\in \widetilde{Y}^{[[s']]}} \mathop{\mathbb{E}}_{\mathbf{m}'\in[-M,M]^{t'}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \Big|\!\Big|\!\Big| \prod_{\underline{\varepsilon}'\in[[s']]} S^{\lfloor A_{\underline{\varepsilon}'}(\mathbf{m}',\mathbf{m})\rfloor + h_{\underline{\varepsilon}'}} F_{\mathbf{m},\mathbf{h}_{\mathbf{m}}} \Big|\!\Big|\!\Big|_{2^{s+1}(\ell+1),S}.$$

Our result will follow if we show that for any integers $h_{\underline{\varepsilon}'}$ we have

$$\limsup_{M\to+\infty} \mathop{\mathbb{E}}_{\mathbf{m}'\in[-M,M]^{t'}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \Big|\!\Big|\!\Big| \prod_{\underline{\varepsilon}'\in[[s']]} S^{\lfloor A_{\underline{\varepsilon}'}(\mathbf{m}',\mathbf{m})\rfloor + h_{\underline{\varepsilon}'}} F_{\mathbf{m},\mathbf{h}_{\mathbf{m}}} \Big|\!\Big|\!\Big|_{2^{s+1}(\ell+1),S} = 0.$$

We substitute $F_{\mathbf{m},\mathbf{h}_{\mathbf{m}}}$ to rewrite this limit as

$$\limsup_{M\to+\infty} \mathop{\mathbb{E}}_{\mathbf{m}'\in[-M,M]^{t'}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \Big|\!\Big|\!\Big| \prod_{\underline{\varepsilon}'\in[[s']]} S^{\lfloor A_{\underline{\varepsilon}'}(\mathbf{m}',\mathbf{m})\rfloor + h_{\underline{\varepsilon}'}} \big(\prod_{\underline{\varepsilon}\in A_1} \mathcal{C}^{|\underline{\varepsilon}|} T^{\lfloor -c_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon},\mathbf{m}}} H_0\big) \Big|\!\Big|\!\Big|_{2^{s+1}(\ell+1),S} =$$
$$\limsup_{M\to+\infty} \mathop{\mathbb{E}}_{\mathbf{m}'\in[-M,M]^{t'}} \mathop{\mathbb{E}}_{\mathbf{m}\in[-M,M]^t} \Big|\!\Big|\!\Big| \prod_{\underline{\varepsilon}'\in[[s']]} \prod_{\underline{\varepsilon}\in A_1} S^{\lfloor A_{\underline{\varepsilon}'}(\mathbf{m}',\mathbf{m})\rfloor + h_{\underline{\varepsilon}'} + \lfloor -c_{\underline{\varepsilon}}(\mathbf{m})\rfloor + h_{\underline{\varepsilon},\mathbf{m}}} (\mathcal{C}^{|\underline{\varepsilon}|} H_0) \Big|\!\Big|\!\Big|_{2^{s+1}(\ell+1),S}. \quad (3.67)$$

For a fixed $\mathbf{m}$ outside all the negligible sets defined above, the polynomials $A_{\underline{\varepsilon}'}(\mathbf{m}', \mathbf{m})$ are pairwise essentially distinct, as polynomials in $\mathbf{m}'$. Therefore, they are also essentially distinct as polynomials in $(\mathbf{m}', \mathbf{m})$. In addition, we have also established that the polynomials $c_{\underline{\varepsilon}}(\mathbf{m})$ are non-constant and essentially distinct. Therefore, it is easy to check that the polynomials $A_{\underline{\varepsilon}'}(\mathbf{m}', \mathbf{m}) - c_{\underline{\varepsilon}}(\mathbf{m})$ are pairwise essentially distinct.

We combine the integer parts in the iterates in (3.67) (correcting with some error terms with values in $\{0, \pm 1\}$). Expanding the seminorm in (3.67), we arrive at an iterated limit of polynomial averages. We also use Lemma 3.2.1 to remove the error terms in the iterates. Using[16] [17, Lemma 4.3], we deduce that the limit in (3.67) is zero under the assumption that $\|\!| H_0 |\!\|_{q,T\times T} = 0$ for some positive integer q. Since

$$\|\!| H_0 |\!\|_{q,T\times T} = \|\!| \overline{f_1} \otimes f_1 |\!\|_{q,T\times T} \leq \|\!| f_1 |\!\|_{q+1,T}^2,$$

we deduce that the desired limit is zero if we assume that $\|\!| f_1 |\!\|_{q+1,T} = 0$. The result follows.

Now that we have established Theorem 1.1.2, we can finish the proof of Theorem 1.1.1.

Proof of Theorem 1.1.1. Note that every 1-good function dominates the logarithmic function $\log t$. Therefore, if the functions $a_1, ..., a_k$ are such that every non-trivial linear combination of them stays logarithmically away from rational polynomials, then the hypotheses of Theorem 1.1.2 are satisfied, which means that the sequences $\lfloor a_1(n) \rfloor, ..., \lfloor a_k(n) \rfloor$ are good for seminorm estimates. Therefore, due to Theorem H we only need to prove that they are good for equidistribution. This, however, follows from the equidistribution results in [5] (i.e. Theorem E) and has been established in [17, Proposition 6.3]. □

[16] This lemma was proven for a specific Følner sequence (namely $[N]^k$), but the same argument extends to the general case. See also [38] for a more detailed proof in the case of integer polynomials.

Chapter 4

Pointwise convergence in nilmanifolds

4.1 Preparations for the proof

In this section, we provide a specific example, which illustrates the central ideas of the proof of Theorem 1.2.2 and skips most of the technical details.

In addition, we will collect some lemmas and make some reductions, which will be useful when we delve into the proof of Theorems 1.2.1 and 1.2.2 in the following sections.

First of all, we present a lemma, which appears in [12, Lemma 3.3] and follows from a simplified version of the arguments of Lemma 3.2.2. We will use this lemma to reduce our problem of studying the long averages over an interval $[1, N]$ (like those appearing in Theorem 1.2.2) to averages in short intervals. Its proof is elementary and so we omit it.

Lemma 4.1.1. *Let $(a(n))_{n\in\mathbb{N}}$ be a bounded sequence of complex numbers. Assume that*

$$\lim_{N\to+\infty} \mathop{\mathbb{E}}_{N\leq n\leq N+L(N)} a(n) = \alpha$$

for some positive function $L(t)$ with $1 \prec L(t) \prec t$. Then, we also have

$$\lim_{N\to+\infty} \mathop{\mathbb{E}}_{1\leq n\leq N} a(n) = \alpha.$$

4.1.1 An example of convergence

Assume $X = G/\Gamma$ is a nilmanifold with G connected and simply connected. We will show that the averages

$$\mathop{\mathbb{E}}_{1\leq n\leq N} f(b_1^{n^{3/2}}x) \cdot g(b_2^{n\log n}x)$$

converge for any $x \in X$, where $b_1, b_2 \in G$.

Using Lemma 4.1.1, it suffices to show that the averages

$$\mathop{\mathbb{E}}_{N\leq n\leq N+L(N)} f(b_1^{n^{3/2}}x) \cdot g(b_2^{n\log n}x)$$

converge, for some sub-linear function $L(t)$. Passing to the nilmanifold $X \times X$, we see that our problem reduces to showing that the averages

$$\mathop{\mathbb{E}}_{N\leq n\leq N+L(N)} F(b_1^{n^{3/2}}b_2^{n\log n}x)$$

converge for any nilmanifold $X = G/\Gamma$, commuting elements[1] $b_1, b_2 \in G$ and function $F \in C(X)$. Due to density, we can actually pick $F \in \mathrm{Lip}(X)$. We provide more details for this deduction in the next section (after Proposition 4.2.1).

[1] When we pass to the product $X \times X$, we have to study the actions of the elements (b_1, e_G) and (e_G, b_2), which clearly commute.

Let X' denote the subnilmanifold $\overline{b_1^{\mathbb{R}} b_2^{\mathbb{R}} \Gamma}$ of X. By Lemma 2.3.2, this set is indeed a subnilmanifold of X and has a representation as H/Δ, with H connected, simply connected and containing all elements b_1^s and b_2^s for any $s \in \mathbb{R}$. In this example, we will also assume that $X' = \overline{b_1^{\mathbb{Z}} b_2^{\mathbb{Z}} \Gamma}$. In the main proof, we will use Lemma 2.3.3 to reduce the general case of the theorem to this one.

Using the Taylor expansion around the point N, we can write

$$(N+h)^{3/2} = N^{3/2} + \frac{3}{2} h N^{1/2} + \frac{3h^2}{8N^{1/2}} - \frac{h^3}{16N^{3/2}} + \frac{3h^4}{128\xi_h^{5/2}}, \quad \text{for some} \quad \xi_h \in [N, N+h]$$

for every $0 \leq h \leq L(N)$. If we choose $L(t)$ to satisfy

$$t^{1/2} \prec L(t) \prec t^{5/8}$$

then the last term in the above expansion is smaller than $o_N(1)$, while the second to last term is unbounded. Similarly, we can write

$$(N+h)\log(N+h) = N \log N + h(\log N + 1) + \frac{h^2}{2N} - \frac{h^3}{6\psi_h^2}, \quad \text{for some} \quad \psi_h \in [N, N+h].$$

If we choose again $L(t)$ to satisfy

$$t^{1/2} \prec L(t) \prec t^{2/3},$$

we can show that the last term is $o_N(1)$, while the h^2 term is unbounded. For instance, we can choose $L(t) = t^{3/5}$ and both growth conditions that we imposed will be satisfied.

Since the function F is continuous, we can disregard the highest order terms in the above expansion since they are both $o_N(1)$. Our problem reduces to showing that the averages

$$\mathop{\mathbb{E}}_{0 \leq h \leq L(N)} F(b_1^{N^{3/2} + \frac{3}{2} h N^{1/2} + \frac{3h^2}{8N^{1/2}} - \frac{h^3}{16N^{3/2}}} b_2^{N \log N + h(\log N + 1) + \frac{h^2}{2N}} x)$$

converge. For the sake of simplicity, we will show that the averages

$$\mathop{\mathbb{E}}_{0 \leq h \leq L(N)} F(b_1^{\frac{h^3}{N^{3/2}}} b_2^{\frac{h^2}{N}} x)$$

converge, since both of these statements follow from the same arguments. For convenience, we will also assume that $x = \Gamma$.

Let $\delta > 0$. We consider the finite sequence

$$(v(h)\Gamma)_{0 \leq h \leq L(N)} = \left(b_1^{\frac{h^3}{N^{3/2}}} b_2^{\frac{h^2}{N}} \Gamma \right)_{0 \leq h \leq L(N)}$$

and we show that, if N is large enough, then it is δ-equidistributed on the subnilmanifold $X' = \overline{b_1^{\mathbb{R}} b_2^{\mathbb{R}} \Gamma}$ of X. It is apparent that $v(n)\Gamma$ is a polynomial sequence in X'. We consider the horizontal torus Z of X', which is isomorphic to some $\mathbb{T}^d$ ($d \in \mathbb{N}$) and we also let π denote the projection map from X' to Z. If the given sequence is not δ-equidistributed (for a fixed value of N), we can invoke Theorem J to find a positive constant $M = M(X', \delta)$ and a non-trivial horizontal character χ_N of modulus at most M and such that

$$\left\| \chi_N(\pi(v(h)\Gamma)) \right\|_{C^\infty[L(N)]} \leq M.$$

Suppose χ_N descends to the character

$$(t_1, ..., t_d) \to e(k_{1,N} t_1 + \cdots + k_{d,N} t_d)$$

on $\mathbb{T}^d$, where $k_{1,N}, ..., k_{d,N}$ are integers. The fact that the modulus is bounded by M implies that

$$|k_{1,N}| + \cdots + |k_{d,N}| \leq M.$$

Let us also write $\pi(b_1\Gamma) = (x_1, ..., x_d)$ and $\pi(b_2\Gamma) = (y_1, ..., y_d)$. Then, the last inequality implies that

$$\|e(\frac{h^3}{N^{3/2}}\sum_{i=1}^{d} k_{i,N}x_i + \frac{h^2}{N}\sum_{i=1}^{d} k_{i,N}y_i)\|_{C^\infty[L(N)]} \leq M. \tag{4.1}$$

Assume there are infinitely many N for which this holds. Since there are only finitely many possible choices for the numbers $k_{1,N}, ..., k_{d,N}$ above, we conclude that there exists a character χ such that $\|\chi(\pi(a(h))\Gamma)\|_{C^\infty[L(N)]} \leq M$ holds for infinitely many $N \in \mathbb{N}$. Then, we rewrite (4.1) (k_i are some integers independent of N) as

$$\|e(\frac{h^3}{N^{3/2}}\sum_{i=1}^{d} k_i x_i + \frac{h^2}{N}\sum_{i=1}^{d} k_i y_i)\|_{C^\infty[L(N)]} \leq M,$$

and this inequality holds for infinitely many N.

The definition of the $C^\infty[L(N)]$ norms implies that we have the relations

$$L(N)^3\|\frac{\sum_{i=1}^{d} k_i x_i}{N^{3/2}}\|_{\mathbb{T}} \leq M$$

and

$$L(N)^2\|\frac{\sum_{i=1}^{d} k_i y_i}{N}\|_{\mathbb{T}} \leq M.$$

for infinitely many N. Due to our choice of the function $L(N)$, these relations fail for N sufficiently large unless

$$\sum_{i=1}^{d} k_i x_i \in \mathbb{Z} \quad \text{and} \quad \sum_{i=1}^{d} k_i y_i \in \mathbb{Z}.$$

This implies that $\chi \circ \pi(b_1\Gamma) = \chi \circ \pi(b_2\Gamma) = 0$ and, consequently, we must also have $\chi \circ \pi(b_1^m b_2^n\Gamma) = 0$ for any $m, n \in \mathbb{Z}$. Since elements of this form are dense in $\overline{b_1^{\mathbb{R}} b_2^{\mathbb{R}}\Gamma}$ by our initial hypothesis, we get that χ must be the trivial character, which is a contradiction.

In conclusion, we have established that the sequence $(v(h)\Gamma)_{0\leq h\leq L(N)}$ is δ-equidistributed for large enough N on $X' = \overline{b_1^{\mathbb{R}} b_2^{\mathbb{R}}\Gamma}$. The result now follows by sending $\delta \to 0$. We also notice that the limit of the averages is $\int_{X'} F\ dm_{X'}$.

Remark. We describe briefly here why we have to use the t^ε term in (1.10) instead of the conjectured optimal term $\log t$. Assuming we had the functions $\log^2 t$ and $t \log t$ in this example, then for any choice of the sub-linear function $L(t)$ that would give a good polynomial approximation for the function $t \log t$, we would have

$$\max_{0\leq h\leq L(N)} |\log^2(N + h) - \log^2 N| = o_N(1),$$

which suggests that the sequence $\log^2 n$ is essentially constant in the small intervals $[N, N + L(N)]$. If we proceed exactly as in the above argument, the best we can actually show is that

$$\Big|\mathop{\mathbb{E}}_{N\leq n\leq N+L(N)} F(b_1^{\log^2 n} b_2^{n\log n}\Gamma) - \int_{Y_2} F(b_1^{\log^2 N} y)\ dm_{Y_2}(y)\Big| \leq \delta \left\|F(b_1^{\log^2 N}\cdot)\right\|_{\mathrm{Lip}(Y_2)}$$

for large enough N, where $Y_2 = \overline{b_2^{\mathbb{R}}\Gamma}$ and $F(b_1^{\log^2 N}\cdot)$ denotes the function $y \to F(b_1^{\log^2 N} y)$ defined on the nilmanifold Y_2. However, the Lipschitz norm above is of the order $\log^2 N\,\|F\|_{\mathrm{Lip}(X)}$, which diverges as $N \to +\infty$, so this bound cannot be useful for any purposes.

Another approach would be to utilize the fact that the parameter M in Theorem J is of the form $\delta^{-O(1)}$, namely we have bounds that are polynomial in δ. Thus, one could allow the parameter δ to vary with N. For instance, establishing a bound of the form $(\log N)^{-(2+\varepsilon)} \left\|F(b_1^{\log^2 N}\cdot)\right\|_{\mathrm{Lip}(Y_2)}$ in place of the term $\delta \left\|F(b_1^{\log^2 N}\cdot)\right\|_{\mathrm{Lip}(Y_2)}$ [2] (namely, showing that our sequence is $(\log N)^{-(2+\varepsilon)}$-equidistributed)

[2] It would actually suffice to obtain this statement for almost all $N \in \mathbb{N}$ in the sense of natural density.

on the right-hand side of the above equation leads to a solution to the more general problem. However, any bound of this type is incorrect in general. Indeed, assume that the horizontal torus of $\overline{b_2^{\mathbb{R}}\Gamma}$ was $\mathbb{T}^2$ and also let $(b_{2,1}, b_{2,2}) \in \mathbb{T}^2$ denote the image of the element $b_2\Gamma$ under the projection map. Following the same approximations as the ones in the example, we would like to show that the finite polynomial sequence $b_2^{h^2/N}\Gamma$, where $0 \leq h \leq L(N)$, is $(\log N)^{-(2+\varepsilon)}$-equidistributed for almost all $N \in \mathbb{N}$ and for some suitable sub-linear function $L(t)$ satisfying only $L(t) \succ t^{1/2}$. Then, an application of Theorem J implies that if this assertion does not hold, then there exists a positive constant C and a horizontal character χ of modulus at most $\log^C N$, such that

$$\left\|\chi(b_2^{h^2/N}\Gamma)\right\|_{C^\infty(L(N))} \leq \log^C N.$$

Equivalently, there exist integers k_1, k_2 with $|k_1| + |k_2| \leq \log^C N$ such that

$$L^2(N)\left\|\frac{k_1 b_{2,1} + k_2 b_{2,2}}{N}\right\|_{\mathbb{T}} \leq \log^C N.$$

Thus, we would get a contradiction if we showed that

$$\min_{|k_1|,|k_2| \leq \log^C N} |k_1 b_{2,1} + k_2 b_{2,2}| \geq \frac{N \log^C N}{L^2(N)}$$

holds for N in a set of density 1. However, we note that bounds like the above depend on the diophantine properties of the numbers $b_{2,1}, b_{2,2}$. Indeed, let us suppose that $\alpha = \frac{b_{2,1}}{b_{2,2}} \leq 1$. If we divide by $b_{2,2}$, the last inequality can be rewritten as

$$\min_{|k_1|,|k_2| \leq \log^C N} \left|k_1\alpha + k_2\right| \geq \frac{N \log^C N}{|b_{2,2}| L^2(N)}.$$

For a fixed choice of k_1, the absolute value is minimized by picking k_2 to be the nearest integer to $-k_1\alpha$. Thus, we would need to show that

$$\min_{|k_1| \leq \log^C N} \left\|k_1\alpha\right\|_{\mathbb{T}} \geq \frac{N \log^C N}{|b_{2,2}| L^2(N)}$$

and we can find $b_{2,1}, b_{2,2} \in (0,1)$ for which this inequality fails for all N in a set of positive upper density. A simpler example that avoids the complicated function on the right-hand side of the last equation is to show that we can find $\alpha \in (0,1)$ for which the inequality $\min_{|k| \leq N} \|k\alpha\|_{\mathbb{T}} \geq 2^{-n}$ fails for all $N \in \mathbb{N}$ in a set of upper density 1. Indeed, we can construct an $\alpha \in (0,1)$ such that $\liminf_{n\to+\infty} 2^{2^n} \|n\alpha\|_{\mathbb{T}} = 0$. Thus, there is a sequence q_n such that $\|q_n a\|_{\mathbb{T}} \leq 2^{-2^{q_n}}$ which implies that $\min_{|k| \leq N} \|k\alpha\|_{\mathbb{T}} \leq 2^{-2^{q_n}} \leq 2^N$ for every N with $q_n \leq N \leq 2^{q_n}$. Thus, the set of N for which the above inequality fails has upper density 1.

4.1.2 Decomposing Hardy field functions

In this section, we will prove a decomposition lemma that is similar to 3.7.1 but which is more suitable for the problem at hand. We consider a Hardy field $\mathcal{H}$ that contains the polynomials and let a be a function in $\mathcal{H}$. We partition $\mathcal{H}$ into equivalence classes by the relation $f \sim g$, which is equivalent to saying that the limit of $f(t)/g(t)$ as $t \to +\infty$ is a non-zero real number. In simple terms, f, g are in the same equivalence class if and only if they have the same growth rate. We put the zero function in its own equivalence class.

We will define the strongly non-polynomial growth rate of a function $a \in \mathcal{H}$ as follows:
i) If a is a strongly non-polynomial function, we define it to be the equivalence class of a.
ii) If a is not strongly non-polynomial, then it can be written in the form $p(t) + x(t)$, where $p(t)$ is a polynomial and $x(t)$ is a strongly non-polynomial function (or the zero function) with $x(t) \prec p(t)$. Observe that $x(t)$ is a function in $\mathcal{H}$, since our Hardy field contains the polynomials. We define the strongly non-polynomial growth rate of a as the equivalence class of the function $x \in \mathcal{H}$.

The strongly non-polynomial growth rate is defined for any function $a \in \mathcal{H}$. It is well defined, in the following sense: consider a function $a \in \mathcal{H}$ like in case ii) above, which has two different representations as $p_1(t) + x_1(t)$ and $p_2(t) + x_2(t)$, where p_1, p_2 are polynomials, x_1, x_2 are strongly non-polynomial and $x_1(t) \prec p_1(t)$ and $x_2(t) \prec p_2(t)$. Then, we must have $x_1(t) \sim x_2(t)$. An example where such distinct representations may exist is the function $a(t) = t^2 + t + t^{3/2}$. We can choose $p_1(t) = t^2, x_1(t) = t^{3/2} + t$ and $p_2(t) = t^2 + t, x_2(t) = t^{3/2}$. While $x_1 \neq x_2$, these two functions have the same growth rate.

A simple observation is that, if a function $a \in \mathcal{H}$ is written in the form $p(t) + x(t)$, where p is polynomial and x is strongly non-polynomial, then the functions a and x have the same strongly non-polynomial growth rate (one could alternatively use this remark to present another equivalent definition of the strongly non-polynomial growth rate).

Finally, we will also say that $a \in \mathcal{H}$ has trivial growth rate, if $\lim\limits_{t\to+\infty} a(t) = 0$. Recall that we also included these functions when we defined the strongly non-polynomial functions.

Now that we are have finished presenting the new terminology, we prove the following lemma.

Lemma 4.1.2. *Let $\mathcal{H}$ be a Hardy field that contains the polynomials and let $a_1, ..., a_k \in \mathcal{H}$ be arbitrary functions. Then, the set $\mathcal{L}(a_1, ..., a_k)$ of non-trivial linear combinations has a basis $(g_1, ..., g_m, h_1, ..., h_\ell)$, where m, ℓ are non-negative integers, such that the functions $h_1, ..., h_\ell$ have the form $p_i(t)+o_t(1)$, where p_i is a real polynomial for every $1 \leq i \leq \ell$ and $g_1, ..., g_m$ have distinct and non-trivial strongly non-polynomial growth rates.*

Proof. We can restrict our attention to the case that the functions $a_1, ..., a_k$ are linearly independent (otherwise, we pass to a maximal subset of these functions whose elements are linearly independent). We induct on k. For $k = 1$, we have nothing to prove. Assume the claim holds for all integers smaller than k. All functions considered below are implicitly assumed to belong to $\mathcal{H}$.

We may write each of the functions $a_1, .., a_k$ in the form $p_i(t) + x_i(t)$ where p_i are real polynomials and $x_i(t)$ are strongly non-polynomial functions (either one of the functions p_i, x_i may also be identically zero). After reordering, we may assume that

$$x_1(t) \gg x_2(t) \gg \cdots \gg x_k(t).$$

Now, we define the number $l \in \{0, 1, ..., k\}$ to be the smallest natural number, for which all functions $x_{l+1}(t), x_{l+2}(t)$ and so on have limit zero (as $t \to +\infty$). If none of the x_i have limits going to 0, then we just set $\ell = k$.

We consider two cases.

i) If the functions $x_1, ..., x_l$ have distinct growth rates, then we are done. In this case, the functions g_j appearing in the statement are the functions $p_i(t) + x_i(t)$ for $1 \leq i \leq l$, while the role of the functions h_j is performed by the functions $p_i(t) + x_i(t)$ for $i > l$ (observe that for $i > l$, we have that $x_i(t)$ have trivial growth rate due to the definition of l). The strongly non-polynomial growth rates of the former set of functions are equal to the growth rates of the functions $x_1, .., x_l$, which are pairwise distinct.

ii) Assume now two of the functions among $x_1, ..., x_l$ have the same growth rate. In particular, let k_0 be the smallest integer such that $x_{k_0} \sim x_{k_0+1}$ (obviously $k_0 < l$) and let $r \geq 1$ be the largest integer such that

$$x_{k_0} \sim x_{k_0+1} \sim \cdots \sim x_{k_0+r}.$$

For $k_0 + 1 \leq i \leq k_0 + r$, we can write $x_i(t) = x_{k_0}(t) + y_i(t)$, where $y_i(t) \prec x_i(t)$. Using this, we can write $a_{k_0}(t) = p_{k_0}(t) + x_{k_0}(t)$ and

$$a_i(t) = (p_{k_0}(t) + x_{k_0}(t)) + (p_i(t) - p_{k_0}(t) + y_i(t)), \quad \text{for } k_0 + 1 \leq i \leq k_0 + r.$$

Now we apply the induction hypothesis on the collection of functions

$$\{p_{k_0+1}(t) - p_{k_0}(t) + y_{k_0+1}(t), ..., p_{k_0+r}(t) - p_{k_0}(t) + y_{k_0+r}(t),$$
$$p_{k_0+r+1}(t) + x_{k_0+r+1}(t), ..., p_k(t) + x_k(t)\}.$$

This gives a basis $(g_1, ..., g_m, u_1, ..., u_\ell)$ for this set of functions, with the properties outlined in the statement. We add the functions $p_1(t) + x_1(t), .., p_{k_0}(t) + x_{k_0}(t)$ to the functions $g_1, ..., g_m$ and add the functions[3] $p_i(t) + x_i(t)$, $l < i \leq k$, to the collection $u_1, ..., u_\ell$. In this way, we construct a basis for the original collection $a_1, ..., a_k$ with the asserted properties (if the functions that we have constructed are not linearly independent, then we can just pass to a subset of these functions that will form a basis). Indeed, we only have to check that the functions

$$p_1(t) + x_1(t), ..., p_{k_0}(t) + x_{k_0}(t), g_1(t), ..., g_m(t)$$

have distinct strongly non-polynomial growth rates. This follows by noting that the strongly non-polynomial growth rates of the functions $g_1, ..., g_m$ cannot be larger than the growth rates of the functions y_i, which all grow strictly slower than x_{k_0}. Thus, the function $p_{k_0}(t) + x_{k_0}(t)$ has bigger strongly non-polynomial growth rate than all of the functions $g_1, ..., g_m$. Furthermore, the strongly non-polynomial growth rate of the function $p_i(t) + x_i(t)$ $(1 \leq i \leq k_0)$ is the same as $x_i(t)$, and these are all pairwise distinct by the definition of k_0. The claim follows. □

Remark. i) Note that we do not require that the functions $a_1, ..., a_k$ have polynomial growth in the above lemma.
ii) A very simple example that illustrates the above decomposition is the following: assume that we have the functions $a_1(t) = t^2 + t^{3/2}, a_2(t) = t^{3/2}, a_3(t) = 2t^{3/2} + t^2$ and $a_4(t) = t^{3/2} + t \log t + t^3$. These four functions are clearly linearly dependent. The above lemma provides the basis (g_1, g_2, h_1), where $g_1(t) = t^{3/2}, g_2(t) = t \log t + t^3$ and $h_1(t) = t^2$. The main property (which will be important in the proof of Theorem 1.2.2) is that the functions g_1, g_2 have distinct strongly non-polynomial growth rates $(t^{3/2}, t \log t$ respectively), even though g_2 grows like t^3 (i.e a polynomial).

4.1.3 Simplifying the assumptions on the nilpotent group

Reduction to connected-simply connected Lie groups

Let G be a k-step nilpotent Lie group and let Γ be a uniform subgroup of G. Then, the space $X = G/\Gamma$ is called a k-step nilmanifold. The space X may have several representations of the form G/Γ (with possible variance in the degree of nilpotency). Let G° be the connected component of e_G in G. If we assume that G/G° is finitely generated[4], then, by passing to the universal cover $\tilde{G}$ of G, it can be shown that X has a representation $\tilde{G}/\tilde{\Gamma}$ where now the underlying group $\tilde{G}$ is simply connected. In addition, we can argue as in [39, Section 1.11] to deduce that X can be embedded as a subnilmanifold in some nilmanifold G'/Γ', where G' is a connected and simply connected nilpotent Lie group and every translation on X has a representation in $X' = G'/\Gamma'$. This means that for any $x \in X$, $b_1, \dots b_k \in G$ and continuous function $F : X \to \mathbb{C}$, we can find $x' \in X'$, $b'_1, \dots, b'_k \in G'$ and $F' : X' \to \mathbb{C}$, such that $F(b_1^{n_1} \dots b_k^{n_k} x) = F'((b'_1)^{n_1} \dots (b'_k)^{n_k} x)$ for all $n_1, \dots, n_k \in \mathbb{Z}$.

Change of base point

For every $b \in G$, we have that the sequence $b^n\Gamma$ is equidistributed in the set $\{\overline{b^n\Gamma \colon n \in \mathbb{Z}}\}$. Therefore, if g is any other element in G, we have that the sequence $b^n g\Gamma$ is equidistributed in the nilmanifold $g\overline{\{(g^{-1}bg)^n\Gamma, n \in \mathbb{N}\}}$. This follows by noting that $b^n g = g(g^{-1}bg)^n$. An analogous relation holds for the elements of the set $(b^s g)_{s\in\mathbb{R}}$, which we define below. This trick, which is called the change of base point trick, can be used when we want to show that some sequence $v(n)x$ is equidistributed (on some specific nilmanifold depending on x) in order to change the base point x to Γ.

4.1.4 Removing the integer parts

In this part, we will establish a lemma that practically implies that part a) of Theorem 1.2.2 follows from part b) of the same theorem. The fact that part a) of Theorem 1.2.1 follows from part b) of the same theorem is precisely the statement of [12, Lemma 5.1], which is proven using very similar

[3]Recall that $x_i(t)$ goes to 0 for $l < i \leq k$.
[4]Without loss of generality we can assume that in our setting, because our results deal with the action of G on finitely many elements of X.

arguments to the proof of Lemma 4.1.3 below. If a collection of sequences of real numbers has the property that the averages

$$\mathop{\mathbb{E}}_{1\leq n\leq N} f_1(b_1^{a_1(n)}x_1)\cdot ...\cdot f_k(b_k^{a_k(n)}x_k) \tag{4.2}$$

converge for all nilmanifolds $X_i = G_i/\Gamma_i$, elements $b_i \in G_i$, points $x_i \in X_i$ and continuous functions f_i defined on X_i, we will say that this collection is pointwise good for nilsystems. The notation $b_i^{a_i(n)}$ makes sense here due to the connectedness assumptions we have imposed on the Lie groups G_i.

Lemma 4.1.3. *Let $a_1(n), ..., a_k(n)$ be sequences of real numbers that satisfy the following:*
a) The collection $a_1(n), ..., a_k(n)$ is pointwise good for nilsystems.
b) For every $1 \leq i \leq k$, we have that the sequence $(a_i(n)\mathbb{Z})_{n\in\mathbb{N}}$ satisfies one of the following:

1. *It is equidistributed on $\mathbb{T}$.*
2. *It converges to some $c = c(i) \in \mathbb{T}$ different from 0.*
3. *It converges to 0 and the sequence $\{a_i(n)\} - \frac{1}{2}$ has a constant sign eventually.*

Then, the sequences $\lfloor a_1(n)\rfloor, ..., \lfloor a_k(n)\rfloor$ are pointwise good for nilsystems.

Remark. The number $\frac{1}{2}$ in the third condition is arbitrary since we could have used any number $\alpha \in (0, 1)$. We primarily use this condition in the following manner: suppose we have a function $f(t)$, which converges monotonically to some $k \in \mathbb{Z}$ as $t \to +\infty$. Then, we clearly have $\|f(t)\|_{\mathbb{T}} \to 0$ and we also observe that the sequence $\{f(n)\}$ does not not oscillate between intervals of the form $[0, \varepsilon]$ and $[1-\varepsilon, 1)$ (due to the monotonicity assumption). Thus, the sequence $\{f(n)\} - \frac{1}{2}$ will indeed have a constant sign (positive if f increases to k and negative otherwise).

Proof. Let $X_i = G_i/\Gamma_i$ be nilmanifolds with G_i connected and simply connected and $b_i \in G_i$. Let $f_1, ..., f_k$ be continuous functions defined on $X_1, ..., X_k$ respectively. Under the hypotheses of the lemma on the sequences $a_1(n), ..., a_k(n)$, we want to show that the averages

$$\mathop{\mathbb{E}}_{1\leq n\leq N} f_1(b_1^{\lfloor a_1(n)\rfloor}x_1)\cdot ...\cdot f_k(b_k^{\lfloor a_k(n)\rfloor}x_k) \tag{4.3}$$

converge for any choice of the $x_i \in X_i$.

Fix some $i \in \{1, 2, ..., k\}$. If the sequence $a_i(n)$ satisfies the second condition, namely that $a_i(n)\mathbb{Z}$ converges to $c\mathbb{Z}$ ($c \neq 0$), then, for n sufficiently large, we have

$$\lfloor a_i(n)\rfloor = a_i(n) - \{c\} + o_n(1).$$

This implies that $b_i^{\lfloor a_i(n)\rfloor} = b_i^{-\{c\}} b_i^{a_i(n)+o_n(1)}$. Since the function f_i is continuous, we can disregard the contribution of the $o_n(1)$ term, while the $b_i^{-\{c\}}$ term can be absorbed by the x_i. Therefore, we notice that in this case, we can remove the integer part for the sequence $a_i(n)$. An entirely similar argument demonstrates that the same holds if $a_i(n)$ satisfies the third condition.

In order to complete the proof, we will consider below the case that each of the sequences $a_i(n)\mathbb{Z}$ is equidistributed on $\mathbb{T}$ for convenience (namely, they all satisfy the first condition). Since we can easily remove the integer parts for those sequences that satisfy the second or third condition as we did above, the argument below easily adapts to the general setting with some changes in notation.

We rewrite the averages in (4.3) as

$$\mathop{\mathbb{E}}_{1\leq n\leq N} \prod_{i=1}^{k} f_i(b_i^{-\{a_i(n)\}} b_i^{a_i(n)} x_i) = \mathop{\mathbb{E}}_{1\leq n\leq N} \prod_{i=1}^{k} \widetilde{f}_i(a_i(n)\mathbb{Z}, b_i^{a_i(n)} x_i)$$

where $\widetilde{f}_i : \mathbb{T} \times X_i \to \mathbb{C}$ is the function defined by the relation

$$\widetilde{f}_i(s\mathbb{Z}, gx) = f_i(b_i^{-\{s\}} gx).$$

Let $v_i(n)$ be the sequence $(a_i(n)\mathbb{Z}, b_i^{a_i(n)}x_i)$. By our hypothesis, for any continuous functions f_i' on $\widetilde{X_i} = \mathbb{T}\times X_i$, the averages of $\prod_{i=1}^k f_i'(v_i(n))$ converge. However, note that the functions $\widetilde{f}_i$ that we are dealing with may have discontinuities when s becomes close to an integer. Our goal is to approximate each $\widetilde{f}_i$ by a continuous function and then use the above observation.

Let $\varepsilon>0$. For every $1\leq i\leq k$, we define a continuous function $f_{i,\varepsilon}$ that agrees everywhere with $\widetilde{f}_i$ on $[\varepsilon, 1-\varepsilon]\times X_i$ and such that $f_{i,\varepsilon}$ is bounded uniformly by $2\left\|\widetilde{f}_i\right\|_\infty$. Observe that

$$\Big|\underset{1\leq n\leq N}{\mathbb{E}}\widetilde{f}_i(v_i(n)) - \underset{1\leq n\leq N}{\mathbb{E}} f_{i,\varepsilon}(v_i(n))\Big| = \frac{1}{N}\Big|\sum_{\substack{1\leq n\leq N\\ a_i(n)\notin[\varepsilon,1-\varepsilon]}}\big(\widetilde{f}_i(v_i(n)) - f_{i,\varepsilon}(v_i(n))\big)\Big| \ll \varepsilon\left\|\widetilde{f}_i\right\|_\infty + o_N(1) \quad (4.4)$$

where the last bound follows from the triangle inequality and the fact that $a_i(n)$ is equidistributed (mod 1), which indicates that the set $\{n\in\mathbb{N}\colon a_i(n)\notin[\varepsilon,1-\varepsilon]\}$ has asymptotic density 2ε.

Combining (4.4) with a simple telescoping argument, we deduce that

$$\limsup_{N\to+\infty}\Big|\underset{1\leq n\leq N}{\mathbb{E}}\prod_{i=1}^k\widetilde{f}_i(v_i(n)) - \underset{1\leq n\leq N}{\mathbb{E}}\prod_{i=1}^k f_{i,\varepsilon}(v_i(n))\Big| \ll k\varepsilon\prod_{i=1}^k\left\|\widetilde{f}_i\right\|_\infty.$$

Since the averages $\underset{1\leq n\leq N}{\mathbb{E}}\prod_{i=1}^k f_{i,\varepsilon}(v(n))$ converge as $N\to\infty$ by our hypothesis (the functions involved here are continuous), we infer that the averages

$$\underset{1\leq n\leq N}{\mathbb{E}}\prod_{i=1}^k\widetilde{f}_i(v_i(n))$$

form a Cauchy sequence and, therefore, converge. The conclusion follows. □

Using the previous lemma, we can establish that the first part of Theorem 1.2.2 follows from the second part. We postpone this until the next section, where we also prove the second part of Theorem 1.2.2.

4.2 Proofs of the pointwise convergence results

The main tool we are going to utilize in our proof is the quantitative Green-Tao theorem on polynomial orbits (Theorem J). A technical obstruction in our proof is that among the functions $a_1,...,a_k$ in the statement of Theorem 1.2.2, we must separate the polynomial functions from the strongly non-polynomial ones. We will accomplish this using Lemma 4.1.2.

First of all, we show that the first part of Theorem 1.2.2 follows from the second part. This is accomplished by using Lemma 4.1.3. We remark again that in part i), there are no connectedness assumptions made on the groups G_i. Nonetheless, our reductions allow us to consider only the case that the Lie groups G_i are connected and simply connected. We implicitly work under this assumption in the proof below.

Proof of part i) of Theorem 1.2.2, assuming part ii). We will have to confirm that the conditions of Lemma 4.1.3 are satisfied. Let $a_1,...,a_k\in\mathcal{H}$ be as in the statement of Theorem 1.2.2. Condition a) of Lemma 4.1.3 is satisfied by our hypothesis. Now, we verify the second condition.

Fix some $i\in\{1,2,..,k\}$. We consider three cases:

i) Assume that the function $a_i(t)$ is such that $|a_i(t)-q(t)|\succ t^\varepsilon$ for all polynomials $q(t)$ with rational coefficients. Then, the sequence $a_i(n)\mathbb{Z}$ is equidistributed on $\mathbb{T}$ (satisfying condition (1)), due to Theorem E.

ii) Assume that the function $a_i(t)$ is such that $\lim_{t\to+\infty}a_i(t)=c\notin\mathbb{Z}$. Then, the sequence $a_i(n)$ satisfies condition (2) of Lemma 4.1.3.

iii) Assume that neither of the above conditions is true. Since $a_i(t)$ must satisfy (1.11), we deduce that $a_i(t)$ converges to some integer c. However, since $a_i(t)$ converges to c monotonically (functions in $\mathcal{H}$ are eventually monotone), we deduce that condition (3) of Lemma 4.1.3 is satisfied and we are done. □

Now we switch our attention to the proof of part ii). Firstly, we will apply Lemma 4.1.2 in order to replace the original functions $a_1, ..., a_k$ with a collection of functions that are more manageable. This will enable us to separate the polynomial functions from strongly non-polynomial ones. In addition, among the strongly non-polynomial functions, we have to isolate those that are sub-fractional, because they behave differently when we try to employ the Taylor expansion. This whole process will reduce Proposition 4.2.1 below to Lemma 4.2.2, which we will then proceed to establish.

Following all these reductions, we use the Taylor expansion to substitute the strongly non-polynomial functions with polynomials in some small intervals. Now, this reduces the original problem to a quantitative equidistribution problem of finite polynomial sequences in a nilmanifold, although the coefficients of the polynomials vary depending on the underlying short interval. Finally, in Step 3, we use the quantitative equidistribution results to show that averages of Lipschitz functions in the nilmanifold over these "variable" polynomial sequences are very close to an integral over a subnilmanifold, which ultimately allows us to evaluate the limit of the initial averages.

We make one final reduction: let $a_1, ..., a_k \in \mathcal{H}$ be functions as in the statement of Theorem 1.2.2. Passing to the product nilmanifold, we infer that our problem follows from the following statement:

Proposition 4.2.1. *Let $X = G/\Gamma$ be a nilmanifold, $b_1, ..., b_k \in G$ are commuting elements and $a_1, ..., a_k \in \mathcal{H}$ have polynomial growth. Assume that there exists $\varepsilon > 0$, such that every function $a \in \mathcal{L}(a_1, ..., a_k)$ satisfies either (1.9) or (1.11). Then, for any $x \in X$ and continuous function $F : X \to \mathbb{C}$, we have that the averages*

$$\mathop{\mathbb{E}}_{1\leq n\leq N} F(b_1^{a_1(n)} \dots b_k^{a_k(n)} x) \tag{4.5}$$

converge.

Proof that Proposition 4.2.1 implies Theorem 1.2.2. We want to show that the averages

$$\frac{1}{N}\sum_{i=1}^{N} f_1(b_1^{a_1(n)} x_1) \cdot ... \cdot f_k(b_k^{a_k(n)} x_k)$$

converge for all $x_i \in X_i$, where the nilmanifolds $X_i = G_i/\Gamma_i$, the elements b_i and the functions $a_i \in \mathcal{H}$ are as in the statement of part (ii) of Theorem 1.2.2. We define the continuous function F on the product nilmanifold $X_1 \times \cdots \times X_k$ by the relation

$$F(y_1, ..., y_k) = f_1(y_1) \cdot ... \cdot f_k(y_k).$$

We also denote by $\widetilde{b_i}$ the element on $G_1 \times \cdots \times G_k$, whose i-th coordinate is equal to b_i, while all of its other coordinates are equal to the respective identity element. Observe that the elements $\widetilde{b}_1, ..., \widetilde{b}_k$ are pairwise commuting. Finally, let us also denote by x the point $(x_1, ..., x_k)$ on the product $X_1 \times \cdots \times X_k$. Then, a simple computation implies that our initial average is equal to

$$\mathop{\mathbb{E}}_{1\leq n\leq N} F(b_1^{a_1(n)} \dots b_k^{a_k(n)} x)$$

and the claim now follows. □

Now, we will reduce Proposition 4.2.1 to the following lemma:

Lemma 4.2.2. *Let G/Γ be a nilmanifold and suppose that $u_1, ..., u_s$ are elements in G, such that*

$$\overline{u_1^{\mathbb{R}} ... u_s^{\mathbb{R}}\Gamma} = \overline{u_1^{\mathbb{Z}} ... u_s^{\mathbb{Z}}\Gamma}. \tag{4.6}$$

In addition, assume that the nilmanifold $X' = \overline{u_1^{\mathbb{R}}...u_s^{\mathbb{R}}\Gamma}$ *can be represented as* G'/Γ'*, where* G' *is connected, simply connected and contains all elements* $u_1, ..., u_s$*. Let* s_0, s *be positive integers and define the sequence* $v(n)$

$$\prod_{i=1}^{s_0} u_i^{p_i(n)+x_i(n)} \prod_{i=s_0+1}^{s} u_i^{\widetilde{p}_i(n)+x_i(n)}, \tag{4.7}$$

where:
a) $p_i, \widetilde{p}_j$ *are polynomials with real coefficients, such that every non-trivial linear combination of the polynomials* $\widetilde{p}_{s_0+1}, ..., \widetilde{p}_s$ *is not an integer polynomial,*
b) the functions x_i *are all strongly non-polynomial, the functions* $x_1, ..., x_{s_0}$ *are not sub-fractional and have pairwise distinct growth rates and the functions* $x_{s_0+1}, ..., x_s$ *are sub-fractional.*
Then, for any Lipschitz function F *on* X' *with Lipschitz norm at most 1, the averages*

$$\underset{1\leq n\leq N}{\mathbb{E}} F(\prod_{i=1}^{s_0} u_i^{p_i(n)+x_i(n)} \prod_{i=s_0+1}^{s} u_i^{\widetilde{p}_i(n)+x_i(n)}\Gamma')$$

converge to the integral $\int_{X'} F \, dm_{X'}$*.*

While the statement may seem relatively convoluted at first, the sequence $v(n)$ above has a convenient form, so that the Taylor approximation can be used directly.

First of all, we prove that Lemma 4.2.2 implies Proposition 4.2.1. We will rely on Lemma 4.1.2 to make the required reductions on the Hardy field functions in the iterates and we will also use Lemma 2.3.3 to get the equality (4.6), where $u_1, \dots, u_s$ will be some appropriate elements of the Lie group G (they will be products of powers of the elements b_i in Proposition 4.2.1).

Proof that Lemma 4.2.2 implies Proposition 4.2.1. Applying Lemma 4.1.2, we can find a basis $f_1, ..., f_s$ for the set $\mathcal{L}(a_1, ..., a_k)$ of non-trivial linear combinations. The collection of functions $f_1, ..., f_s$ can be written in the form $(g_1, , , ., g_m, h_1, ..., h_\ell)$ where g_i, h_i are as in Lemma 4.1.2. We will not use this specific property until a little further below, so as to avoid cumbersome notation. Note that the fact that $f_1, ..., f_s$ form a basis indicates that the assumptions on the linear combinations of the $a_1, ..., a_k$ in the statement of Proposition 4.2.1 are now transferred to the functions $f_1, ..., f_s$.

If we write

$$a_i(t) = \sum_{j=1}^{s} c_{i,j} f_j(t), \tag{4.8}$$

for some real numbers $c_{i,j}$, then we can rewrite the average in (4.5) as

$$\underset{1\leq n\leq N}{\mathbb{E}} F(u_1^{f_1(n)} \dots u_s^{f_s(n)} x) \tag{4.9}$$

for some commuting elements $u_1, ..., u_s \in G$ (here, the fact that the elements $b_1, ..., b_k$ commute is required). We denote

$$v(n) = u_1^{f_1(n)} \dots u_s^{f_s(n)},$$

which is a sequence in G. We want to establish that the averages of the sequence $F(v(n)x)$ converge for all $x \in X$ and any continuous function F. If one of the functions $f_1, ..., f_m$ is such that the limit $\lim\limits_{t\to+\infty} f_i(t)$ is a real number (which can be the case when a linear combination of the original functions satisfies (1.11)), we can invoke the continuity of F to eliminate the corresponding term $u_i^{f_i(n)}$ in the product and replace it by a constant. Hence, we may assume that all of the functions $f_1(t), ..., f_s(t)$ go to $\pm\infty$, as $t \to +\infty$.

Now we use the particular structure of the functions $f_1, ..., f_s$. The statement of Lemma 4.1.2 implies that the collection of functions $f_1, ..., f_s$ has the form $(g_1, ..., g_m, h_1, ..., h_\ell)$ (clearly, $m + \ell = s$) such that the functions g_i can be written in the form $p_i(t)+x_i(t)$, where the functions $x_1(t), ..., x_m(t)$ are strongly non-polynomial and have pairwise distinct (and non-trivial) growth rates, while the functions h_i can be written in the form $\widetilde{p}_i(t) + y_i(t)$, where $y_i(t)$ converges to 0. Here, p_i and $\widetilde{p}_i$ are polynomials with real coefficients.

We may rearrange the functions f_i so that $f_i = g_i$ for all $1 \leq i \leq m$ and $f_j = h_{j-m}$ for each $m+1 \leq j \leq s$. Rewrite the sequence $v(n)$ as

$$v(n) = \prod_{i=1}^{m} u_i^{g_i(n)} \cdot \prod_{i=1}^{\ell} u_{m+i}^{h_i(n)} = \prod_{i=1}^{m} u_i^{p_i(n)+x_i(n)} \cdot \prod_{i=1}^{\ell} w_i^{\widetilde{p}_i(n)+y_i(n)},$$

where we use the notation w_i for the element u_{i+m} in the last equality. Without loss of generality, assume that

$$x_1(t) \succ x_2(t) \succ \cdots \succ x_m(t) \succ 1.$$

Firstly, we need to distinguish between the sub-fractional functions and the "fast" growing functions among the functions $x_i(t)$ (this will be important later when we use the polynomial expansion). Thus, let $0 \leq s_0 \leq m$ be a natural number such that $x_{s_0}(t) \gg t^{\varepsilon}$ for some $\varepsilon > 0$, while x_{s_0+1} is a sub-fractional function. This also implies that all the functions x_i for i satisfying $s_0 + 1 \leq i \leq m$ are sub-fractional since we have arranged the functions so that their growth rates are in descending order.

Once again, we rewrite the sequence $v(n)$ in the form

$$v(n) = \prod_{i=1}^{s_0} u_i^{p_i(n)+x_i(n)} \prod_{i=s_0+1}^{m} u_i^{p_i(n)+x_i(n)} \prod_{i=1}^{\ell} w_i^{\widetilde{p}_i(n)+y_i(n)}.$$

Because the function F is continuous, we can discard the functions $y_1, ..., y_\ell$, since they all converge to zero. The hypotheses (1.10) and (1.11) on the linear combinations of the remaining functions in the exponents continue to hold. Indeed, this can be seen by noting that (1.10) and (1.11) still hold when replacing one of the functions (say a_1) by a function of the form $a_1(t) + e(t)$, with $e(t) \to 0$. Consequently, we can redefine $v(n)$ to be the sequence

$$v(n) = \prod_{i=1}^{s_0} u_i^{p_i(n)+x_i(n)} \prod_{i=s_0+1}^{m} u_i^{p_i(n)+x_i(n)} \prod_{i=1}^{\ell} w_i^{\widetilde{p}_i(n)}.$$

We will now reduce our problem to the case that the polynomials $\widetilde{p}_1(t), ..., \widetilde{p}_\ell(t)$ are linearly independent. Due to our hypothesis (namely (1.10),(1.11)), every non-trivial linear combination of the functions $\widetilde{p}_1(t), ..., \widetilde{p}_\ell(t)$ must satisfy either (1.10) or (1.11). Thus, every linear combination of the polynomials $\widetilde{p}_1(t), ..., \widetilde{p}_\ell(t)$ is not a polynomial with integer coefficients unless it is the zero polynomial. If the second case is true, there exist $c_1, ..., c_{\ell-1} \in \mathbb{R}$ such that

$$\widetilde{p}_\ell = c_1\widetilde{p}_1 + \cdots + c_{\ell-1}\widetilde{p}_{\ell-1}.$$

Then, we have

$$\prod_{i=1}^{\ell} w_i^{\widetilde{p}_i(n)} = \prod_{i=1}^{\ell-1} (w_i w_\ell^{c_i})^{\widetilde{p}_i(n)}.$$

If the polynomials $\widetilde{p}_1, ..., \widetilde{p}_{\ell-1}$ are linearly independent, then we are done. Otherwise, we proceed similarly to eliminate $\widetilde{p}_{\ell-1}$. After a finite number of steps, we will reach a collection of linearly independent polynomials.

In view of the above, we are allowed to assume that $\widetilde{p}_1, ..., \widetilde{p}_\ell$ are linearly independent. Now, we show that we can reduce to the case that the polynomials $p_{s_0+1}, \ldots, p_m, \widetilde{p}_1, \ldots, \widetilde{p}_\ell$. Indeed, the linear independence assumption on the polynomials $\widetilde{p}_1, \ldots, \widetilde{p}_\ell$ implies that the polynomials $p_{s_0+1}, ..., p_m, \widetilde{p}_1, ..., \widetilde{p}_\ell$ are linearly independent. To see how this works, observe that if there are real numbers c_i, d_j such that

$$\sum_{i=1}^{m-s_0} c_i p_{s_0+i} + \sum_{i=1}^{\ell} d_i p'_i = 0,$$

then the function

$$\sum_{i=1}^{m-s_0} c_i(p_{s_0+i} + x_{s_0+i}) + \sum_{i=1}^{\ell} d_i\widetilde{p}_i = \sum_{i=1}^{m-s_0} c_i x_{s_0+i}$$

is a sub-fractional function that does not converge to 0, since the functions x_{s_0+i} are sub-fractional and have pairwise distinct growth rates. This contradicts our hypothesis (specifically (1.11)) and our claim follows.

In conclusion, we see that the sequence $v(n)$ can be written in the form

$$\prod_{i=1}^{s_0} u_i^{p_i(n)+x_i(n)} \prod_{i=s_0+1}^{m} u_i^{p_i(n)+x_i(n)} \prod_{i=1}^{\ell} w_i^{\widetilde{p}_i(n)}, \tag{4.10}$$

where the functions x_i are strongly non-polynomial with distinct growth rates, the functions $x_1, ..., x_{s_0}$ are not sub-fractional, the functions $x_{s_0+1}, ..., x_s$ are sub-fractional and every non-trivial linear combination of the polynomials $p_{s_0+1}, ..., p_m, \widetilde{p}_1, ..., \widetilde{p}_\ell$ is not an integer polynomial. We also recall that we have arranged the functions x_i to be in decreasing order with respect to their growth rates.

We can combine the last two factors of this product into one factor to simplify our problem a bit more. More specifically, we can rewrite the sequence $v(n)$ in the form (we make some mild modifications in our notation here)

$$v(n) = \prod_{i=1}^{s_0} u_i^{p_i(n)+x_i(n)} \prod_{i=s_0+1}^{s} u_i^{\widetilde{p}_i(n)+x_i(n)}, \tag{4.11}$$

where $s = m + l$, $p_i, \widetilde{p}_j$ are real polynomials, the functions x_i are strongly non-polynomial with distinct growth rates, $x_1, ..., x_{s_0}$ are not sub-fractional, $x_{s_0+1}, ..., x_s$ are sub-fractional and every non-trivial linear combination of the polynomials $\widetilde{p}_i$ is not an integer polynomial. Namely, our functions satisfy hypotheses a) and b) of Lemma 4.2.2.

In order to establish our assertion, it suffices to show that the sequence $v(n)x$ (where $v(n)$ is as in (4.11)) is equidistributed on the nilmanifold $X' = \overline{u_1^{\mathbb{R}} \dots u_s^{\mathbb{R}} x}$ for any $x \in X$. We will prove this in the case $x = \Gamma$ since the general case follows from this using the change of base point trick (see Subsection 4.1.3). In addition, we can invoke Lemma 2.3.3 to find a real number s_0, such that $X' = \overline{(u_1^{s_0})^{\mathbb{Z}} ... (u_s^{s_0})^{\mathbb{Z}} \Gamma}$. Replacing the functions $p_i(t) + x_i(t)$ $(1 \leq i \leq s_0)$ by the functions $\big(p_i(t) + x_i(t)\big)/s_0$ and $\widetilde{p}_i(t) + x_i(t)$ $(s_0 + 1 \leq i \leq s)$ by $\big(\widetilde{p}_i(t) + x_i(t)\big)/s_0$ (the assumptions on the linear combinations of the functions remain unaffected), we can reduce our problem to the case that $X' = \overline{u_1^{\mathbb{Z}} ... u_s^{\mathbb{Z}} \Gamma}$.

We want to show that for any continuous function F from $X' = G'/\Gamma'$ (G' is connected, simply connected and Γ' is a uniform subgroup), the averages

$$\mathop{\mathbb{E}}_{1 \leq n \leq N} F(v(n)\Gamma')$$

converge to the integral $\int_{X'} F \, dm_{X'}$. Since Lipschitz functions are dense in the space $C(X')$, we may assume that F is Lipschitz continuous. In addition, we may assume after rescaling that $\|F\|_{\mathrm{Lip}(X')} \leq 1$. Now, our claim follows immediately from Lemma 4.2.2. □

In the following part, we will prove Lemma 4.2.2. We split the proof into two steps. During Step 1, we will approximate the functions $x_1, \dots, x_s$ by polynomials in a suitable short interval. Our goal is to reach an average over a short interval of the form $[N, N + L(N)]$ of a sequence of the form $F(g(n)x)$, where F is Lipschitz and $g(n)$ is a polynomial sequence on the nilmanifold X' (the polynomial sequence will vary with the parameter N). This will be ensured by the results we already have on the simultaneous Taylor expansion of Hardy field functions. In step 2, we will use Theorem J to deduce that these averages are close to the integral of F for large values of N.

All the reductions above allow us to write $v(n)$ in a form that will be appropriate for the application of the quantitative equidistribution theorem (after we perform the Taylor expansion). When we apply the Taylor expansion in the first step, the functions $x_{s_0+1}, ..., x_s$ will become approximately constant and thus the desired equidistribution will be mainly "affected" by the polynomials $\widetilde{p}_{s_0+1}, ..., \widetilde{p}_s$. On the other hand, the functions $x_1, ..., x_{s_0}$ will play a meaningful role in the equidistribution of our sequence. In particular, the presence of the functions $x_1, \dots, x_{s_0}$ will imply "closeness" of our averages to the integral of the Lipschitz function F, unless the projections of the elements $u_1, \dots, u_{s_0}$ on the horizontal torus are zero. In this second case, condition $a)$ on the polynomials completes the proof. Lastly, the "linear independence" condition of the polynomials $\widetilde{p}_{s_0+1},...,\widetilde{p}_s$ guarantees that the projection of the sequence $v(n)$ on X' will be equidistributed on the entire nilmanifold $\overline{u_1^{\mathbb{R}}...u_s^{\mathbb{R}}\Gamma}$, since, otherwise, we would need to pass to some subnilmanifold to guarantee equidistribution (and to an appropriate arithmetic progression).

Proof of Lemma 4.2.2. **Step 1: Approximating by polynomials:** Let $L(t)$ be a sub-linear function with $\lim_{t\to+\infty} L(t) = +\infty$ that we will determine later. It suffices to show that the sequence of the averages

$$\mathop{\mathbb{E}}_{N\leq n\leq N+L(N)} F(v(n)\Gamma') \tag{4.12}$$

converges to $\int_{X'} F\, dm_{X'}$, since the conclusion would follow from Lemma 4.1.1. Reordering if necessary, we assume again that

$$x_1(t) \succ \cdots \succ x_{s_0}(t).$$

Let r be a very large natural number compared to the degrees of the polynomials $p_i, \widetilde{p}_j$ and the degrees of the functions $x_i(t)$. If r is sufficiently large, we have that $x_i^{(r)}(t) = o_t(1)$ for all $i \in \{1,\ldots,s_0\}$. Assuming again that r is sufficiently large, then for any function $L(t)$ that satisfies

$$(x_i^{(r)}(t))^{-1/r} \prec L(t) \prec t^{1-\varepsilon'}$$

for some $\varepsilon' > 0$ and all $i \in 1,\ldots,s_0$, we have that for each $i \in \{1,...,s_0\}$, there is a unique natural number $k_i \geq r$ so that the sub-class $S(x_i,k_i)$ contains the function $L(t)$ (this follows from Lemma 3.5.1). The fact that the function $L(t)$ belongs to $S(x_i,k_i)$ indicates that we have the relations

$$(x_i^{(k_i)}(t))^{-1/k_i} \prec L(t) \prec (x_i^{(k_i+1)}(t))^{-1/(k_i+1)}. \tag{4.13}$$

We can guarantee that the numbers k_i are also very large compared to the degrees of the polynomials $p_j, \widetilde{p}_{j'}$ by enlarging the number r in the beginning[5].

We use the Taylor expansion for the functions $x_1(t),...,x_{s_0}(t)$ to write

$$x_i(N+h) = x_i(N) + \cdots + \frac{x_i^{k_i}(N)h^k}{k_i!} + o_N(1) = q_{i,N}(h) + o_N(1) \tag{4.14}$$

for $0 \leq h \leq L(N)$. If, on the other hand, we have $i > s_0$ (namely, in the case where the function x_i is sub-fractional), then

$$\max_{0\leq h\leq L(N)} |x_i(N+h) - x_i(N)| = o_N(1). \tag{4.15}$$

In addition, we denote $p_{i,N}(h) = p_i(N+h)$ and similarly $\widetilde{p}_{i,N}(h) = \widetilde{p}_i(N+h)$ for every admissible value of i. Thus, we rewrite the expression in (4.12) as

$$\mathop{\mathbb{E}}_{0\leq h\leq L(N)} F(w_N \prod_{i=1}^{s_0} u_i^{q_{i,N}(h)+p_{i,N}(h)} \prod_{i=s_0+1}^{s} u_i^{\widetilde{p}_{i,N}(h)}\Gamma') \tag{4.16}$$

where we discarded the $o_N(1)$ terms, because F is continuous. Here, $w_N = \prod_{i=s_0+1}^{s} u_i^{x_i(N)}$ but the explicit form of this term will not concern us, since we will only require that the element w_N belongs to the underlying group G' defining the nilmanifold $X' = \overline{u_1^{\mathbb{R}}\ldots u_s^{\mathbb{R}}\Gamma}$

In conclusion, we have reduced our problem to showing that given the nilmanifold $X' = \overline{u_1^{\mathbb{R}}\ldots u_s^{\mathbb{R}}\Gamma}$ (which is also equal to $\overline{u_1^{\mathbb{Z}}\ldots u_s^{\mathbb{Z}}\Gamma}$), the averages in (4.16) converge. Here, the polynomials $q_{i,N}$ are defined in (4.14) (they are essentially the Taylor polynomials of the Hardy field functions x_i), while the polynomials $p_{i,N}, \widetilde{p}_{j,N}$ were defined by the relations $p_{i,N} = p_i(N+h)$ and $\widetilde{p}_{j,N} = \widetilde{p}_j(N+h)$, where the $p_i, \widetilde{p}_j$ are polynomials with real coefficients. We also recall that the polynomials $\widetilde{p}_i$ are such that every non-trivial linear combination of them is not an integer polynomial. Under all these assumptions, we

[5]For example, assuming that k_i is at least 10 times as large as the maximal degree appearing among the polynomials $p_i, \widetilde{p}_j$ and 10 times as large as the number s of all existing polynomials would suffice for our arguments.

will show that the polynomial sequence (restricted to the range $0 \le h \le L(N)$) inside the function F is δ-equidistributed for N sufficiently large in the following step. We remark that the growth conditions (4.13) imposed on the function $L(t)$ will also play a crucial role in this.

Step 2: Using the quantitative equidistribution theorem: Let $Z \cong \mathbb{T}^d$ be the horizontal torus of the nilmanifold $X' = \overline{u_1^{\mathbb{R}} \dots u_s^{\mathbb{R}}\Gamma}$ and let $\pi : X' \to Z$ denote the projection map. Let $\delta > 0$ be sufficiently small (in the sense that Theorem J is applicable). We assert that the finite polynomial sequence

$$\Big(\prod_{i=1}^{s_0} u_i^{q_{i,N}(h)+p_{i,N}(h)} \prod_{i=s_0+1}^{s} u_i^{\widetilde{p}_{i,N}(h)}\Gamma'\Big)_{0\le h\le L(N)} \tag{4.17}$$

is δ-equidistributed on the nilmanifold X' for N sufficiently large. If the claim does not hold for a natural number N, then by Theorem J, there exists a real number[6] $M > 0$ and a non-trivial horizontal character χ_N of modulus $\le M$ such that

$$\|\chi_N \circ \pi(\prod_{i=1}^{s_0} u_i^{q_{i,N}(h)+p_{i,N}(h)} \prod_{i=s_0+1}^{s} u_i^{\widetilde{p}_{i,N}(h)}\Gamma')\|_{C^\infty[L(N)]} \le M. \tag{4.18}$$

Thus, if our prior assertion fails, then the above relation would hold for infinitely many $N \in \mathbb{N}$.

Our first goal is to eliminate the dependence of the characters χ_N on the variable N. Note that the function $\chi_N \circ \pi$ is a character on $\mathbb{T}^d$ of modulus $\le M$ and, thus, has the form

$$(t_1, ..., t_d) \to e(k_{1,N}t_1 + \dots + k_{d,N}t_d)$$

for $k_{i,N} \in \mathbb{Z}$ with $|k_{1,N}| + \dots + |k_{d,N}| \le M$. We also write $\pi(u_i) = (u_{i,1}, ..., u_{i,d})$ for the projections of the elements u_i on the horizontal torus. Then, a straightforward computation allows us to rewrite (4.18) as

$$\|e(\sum_{i=1}^{s_0}(q_{i,N}(h) + p_{i,N}(h))(k_{1,N}u_{i,1} + \dots + k_{d,N}u_{i,d}) + \sum_{i=s_0+1}^{s} (\widetilde{p}_{i,N}(h))(k_{1,N}u_{i,1} + \dots + k_{d,N}u_{i,d}))\|_{C^\infty[L(N)]} \le M. \tag{4.19}$$

Since there are only finitely many choices for the numbers $k_{1,N}, ..., k_{d,N}$, we have that, if our claim fails, there are $k_1, ..., k_d \in \mathbb{Z}$, so that the inequality

$$\|e(\sum_{i=1}^{s_0}(q_{i,N}(h) + p_{i,N}(h))(k_1u_{i,1} + \dots + k_du_{i,d}) + \sum_{i=s_0+1}^{s} (\widetilde{p}_{i,N}(h))(k_1u_{i,1} + \dots + k_du_{i,d}))\|_{C^\infty[L(N)]} \le M. \tag{4.20}$$

holds for infinitely many $N \in \mathbb{N}$. We will also denote the horizontal character corresponding to the d-tuplet $(k_1, ..., k_d)$ by χ. Thus, we have eliminated the dependence of the character χ on N.

Denote $\widetilde{u_i} = k_1u_{i,1} + \dots + k_du_{i,d}$. We will show that the above hypotheses imply that all the numbers $\widetilde{u_i}$ equal 0. Thus, suppose that this is not valid and we will reach a contradiction. We consider two cases:

Case 1: Firstly, suppose that all of the numbers $\widetilde{u_i}$ with $1 \le i \le s_0$ are zero, which implies that the first summand in (4.20) vanishes. Naturally, (4.20) is simplified to

$$\|e(\sum_{i=s_0+1}^{s} \widetilde{p}_{i,N}(h)\widetilde{u_i})\|_{C^\infty[L(N)]} \le M. \tag{4.21}$$

[6]The constant M depends only on δ, the nilmanifold X' as well as the degrees of the polynomials q_i, p_i, which are all fixed in our arguments. The central property we need is that it is independent of the variable N.

We recall here that we had defined $\widetilde{p}_{i,N}(h) = \widetilde{p}_i(N+h)$. Let $Q(t) = \sum_{i=s_0+1}^{s} \widetilde{u}_i\widetilde{p}_i(t)$. This is a linear combination of the polynomials $\widetilde{p}_i(t)$. However, this linear combination is not a polynomial in $\mathbb{Q}[t]$ due to our assumptions on the polynomials $\widetilde{p}_i(n)$, unless, of course, all the coefficients $\widetilde{u}_i$ (for $s_0+1 \leq i \leq s$) in this combination are zero, which we have supposed to not be the case. Thus, $Q(t)$ has at least one irrational coefficient (except the constant term) and is equidistributed on $\mathbb{T}$. The relation (4.21) implies that $\|e(Q(N+h))\|_{C^\infty[L(N)]} \leq M$ for infinitely many N. It is not difficult to see by calculating the coefficients in $Q(N+h)$ that this fails for N large enough.

Case 2: Suppose now that at least one of the numbers $\widetilde{u}_i$ with $1 \leq i \leq s_0$ is non-zero Furthermore, assume l is a positive integer that is larger than the degrees of the polynomials $p_{i,N}(h), \widetilde{p}_{j,N}(h)$ (for all admissible values of the indices i,j) as well as the degrees of the functions x_i, but l is also smaller than all the numbers k_i. Recall that we have picked k_i to be very large in relation to the degrees of the polynomials $p_i, \widetilde{p}_j$ and degrees of the functions x_i in the beginning, thus we can find "many" such numbers l. The fact that l is larger than the degrees of the functions x_i combined with Lemma 2.1.3 implies that $x_i^{(l)}(t) \to 0$, as $t \to +\infty$.

For a number l as above the coefficient of the term h^l in the polynomial appearing in (4.20) is equal to

$$\frac{1}{l!}\sum_{i=1}^{s_0} x_i^{(l)}(N)\widetilde{u}_i$$

and, thus, it does not depend on the polynomials $p_i, \widetilde{p}_j$. Using the definition of the smoothness norms, (4.20) implies that

$$L(N)^l \Big\| \frac{1}{l!}\sum_{i=1}^{s_0} x_i^{(l)}(N)\widetilde{u}_i \Big\|_{\mathbb{T}} \leq M$$

for infinitely many $N \in \mathbb{N}$. The last inequality becomes

$$L(N)^l \Big| \sum_{i=1}^{s_0} x_i^{(l)}(N)\widetilde{u}_i \Big| \leq l!M,$$

for large enough N, because all functions $x_i^l(t)$ go to 0. However, the Hardy field function inside the absolute value above has the same growth rate as the function $x_1^{(l)}(t)$, since the functions $x_1, ..., x_{s_0}$ are strongly non-polynomial and have distinct growth rates (recall that x_1 has the largest growth rate among the x_i), unless, of course, $\widetilde{u}_1 = 0$. If the latter does not hold, we get

$$|x_1^{(l)}(N)\widetilde{u}_1| \leq \frac{C}{L(N)^l}$$

for infinitely many N and some constant C, which contradicts (4.13). Thus, we eventually deduce that $\widetilde{u}_1 = k_1 u_{1,1} + \cdots + k_d u_{1,d} = 0$. Repeating the same argument, we get inductively that $\widetilde{u}_i = k_1 u_{i,1} + \cdots + k_d u_{i,d}{=}0$ for all $1 \leq i \leq s_0$, which is a contradiction.

To summarize, we have shown that if the sequence in (4.17) is not δ-equidistributed for all large enough N, then all the numbers $\widetilde{u}_i = k_1 u_{i,1} + \cdots + k_d u_{i,d}$ are zero. Equivalently, we have $\chi \circ \pi(u_i) = 0$ for all $1 \leq i \leq s$. This implies that the character χ is the trivial character on X'. Indeed, the character χ annihilates all elements $u_1^{n_1} \cdots u_s^{n_s}\Gamma$, where $n_1, ..., n_s \in \mathbb{Z}$ and by density of those elements on X' (recall our assumption that X' is also equal to the nilmanifold $\overline{u_1^{\mathbb{Z}} ... u_s^{\mathbb{Z}}\Gamma}$), χ is zero everywhere. This is a contradiction (the horizontal characters appearing when we applied Theorem J are assumed to be non-trivial).

In conclusion, we have that the finite polynomial sequence in (4.17) is δ-equidistributed for N sufficiently large. Thus, we conclude that the averages in (4.12) are $\delta \left\|F(w_N\cdot)\right\|_{\mathrm{Lip}(X')} = \delta \left\|F\right\|_{\mathrm{Lip}(X')}$ close to the quantity $\int_{X'} F(w_N x)\, dm_{X'}(x)$. The action of w_N on X' preserves the Haar measure of X', so we get that the last integral is equal to $\int_{X'} F(x)\, dm_{X'}(x)$. Taking $\delta \to 0$, we finish the proof. □

Proof of Theorem 1.2.1. As we explained in the previous section (before the statement of Lemma 4.1.3), the first part follows from the second part (see also [12, Lemma 5.1]) and, in turn, this second part follows using similar arguments as in the proof of Theorem 1.2.2. We only highlight the main differences here. All the disparities appear in the part where we reduce Proposition 4.2.1 to Lemma 4.2.2.

a) In (4.9), all the functions $f_1, ..., f_s$ satisfy (1.10) (there are no functions among the f_i that satisfy $\lim\limits_{t\to+\infty} |f_i(t)| < \infty$). We also have $k = s$.

b) We do not have to make the reduction to the case where the polynomials $\widetilde{p}_1, ..., \widetilde{p}_\ell$ are linearly independent. There cannot be a non-trivial linear combination of them that is zero, because that would violate (1.10).

c) The limit of the averages is again $\int_{X'} F(x)\, dm_{X'}(x)$, where $X' = \overline{u_1^{\mathbb{R}}...u_s^{\mathbb{R}}\Gamma}$ by Lemma 4.2.2. We would like to show that the limit is equal to $\int_{X''} F dm''_X$, where X'' is the nilmanifold $\overline{b_1^{\mathbb{R}}...b_k^{\mathbb{R}}\Gamma}$. Recall that each u_i is equal to $b_1^{c_{i,1}}...b_k^{c_{i,k}}$ (by (4.8)) and the numbers $c_{i,j}$ form an invertible $k\times k$ matrix (due to the linear independence assumption on the original functions $a_1, ..., a_k$). Thus, we can also write $b_i = \prod_{j=1}^k u_i^{c'_{i,j}}$ for some numbers $c'_{i,j}$ (here, we also use that the elements b_i are pairwise commuting). Combining the above, we have that $b_1^{\mathbb{R}}...b_k^{\mathbb{R}} = u_1^{\mathbb{R}}...u_k^{\mathbb{R}}$ and thus the closures of their projections on G/Γ define the same subnilmanifold. □

4.2.1 Convergence for linearly dependent iterates

Finally, we provide a proof of Theorem 1.2.3. Here, we will use Theorem 1.1.2. Although it will not be used in the proof, we have to assume below that the Hardy field $\mathcal{H}$ that we work with is closed under composition and compositional inversion of functions, since the seminorm estimates were proven under this assumption in the previous chapter.

Proof. Using a standard ergodic decomposition argument, we may assume that the system (X, μ, T) is ergodic. We can also rescale the functions $f_i \in L^\infty(\mu)$ so that they are 1-bounded. Our first objective is to apply Theorem 1.1.2 (or rather, the more flexible Proposition 3.1.1), in order to reduce the problem to the case where the system X is a nilsystem. If the functions $a_1, ..., a_k$ are such that the conditions of Proposition 3.1.1 are satisfied, then this can be done instantly. If this does not hold, we have to perform a series of reductions to be able to apply Proposition 3.1.1. We do this in 2 steps:

a) Firstly, assume there exists one function among the a_i (say a_1 for simplicity), which has growth rate smaller than or equal to $\log t$. Then, using (1.10) and (1.11), we deduce that a_1 converges monotonically to some real number c and the integer part of $a_1(n)$ becomes a constant. Thus, the asymptotic behavior of the averages in (1.12) is the same, if we substitute the term $T^{\lfloor a_1(n)\rfloor f_1}$ with the term $T^{\lfloor c\rfloor} f_1$. Consequently, we only need to show that the averages

$$\mathop{\mathbb{E}}_{1\leq n\leq N} T^{\lfloor a_2(n)\rfloor} f_2 \dots T^{\lfloor a_k(n)\rfloor} f_k$$

converge in norm. Repeating the same argument, we eliminate all functions a_i that grow slower than $\log t$.

b) Due to the reduction in the previous step, we have a sub-collection of the original functions, so that all functions in this new set dominate $\log t$. We will denote this collection by $a_1, ..., a_k$ again, and our task is to show that the averages

$$\mathop{\mathbb{E}}_{1\leq n\leq N} T^{\lfloor a_1(n)\rfloor} f_1 \dots T^{\lfloor a_k(n)\rfloor} f_k$$

converge in mean (for all systems). Our next objective is to eliminate pairs of functions, whose difference grows slower than $\log t$ so that we can ultimately apply Proposition 3.1.1.

Assume that two of the functions (say a_1, a_2) are such that their difference is dominated by $\log t$. We observe that the function $a_1(t)$ goes to $\pm\infty$ as $t \to +\infty$, since it dominates $\log t$. In that case, the function $a_1(t)$ satisfies (1.10) and by Theorem E, the sequence $a_1(n)$ is equidistributed (mod 1). Observe that since $a_1 - a_2$ must satisfy (1.11), we must have $a_2(t) = a_1(t) + c + x(t)$, where the function $x(t) \in \mathcal{H}$ converges to 0 monotonically and c is a real number. Thus, for $t \in \mathbb{R}$ sufficiently large we have

$$\lfloor a_2(t)\rfloor = \lfloor a_1(t) + x(t) + c\rfloor = \lfloor a_1(t)\rfloor + \lfloor c\rfloor + \varepsilon(t) \tag{4.22}$$

where $\varepsilon(t) \in \{0, \pm 1, \pm 2\}$ and the value of $\varepsilon(t)$ depends on whether the inequalities

$$\{a_1(t) + c\} + \{x(t)\} \leq 1$$

and

$$\{a_1(t)\} + \{c\} \leq 1$$

hold or not, as well as whether $x(t)$ is eventually positive or negative.

Define $A_z = \{t \in \mathbb{R}, \varepsilon(t) = z\}$ for $z \in \{0, \pm 1, \pm 2\}$. Then, we see that our multiple averages are equal to the sum

$$\sum_{z \in \{0, \pm 1, \pm 2\}} \mathop{\mathbb{E}}_{1 \leq n \leq N} \mathbf{1}_{A_z}(n)\, T^{\lfloor a_1(n) \rfloor}(f_1 \cdot T^{\lfloor c \rfloor + z} f_2) \cdot T^{\lfloor a_3(n) \rfloor} f_3 \dots T^{\lfloor a_k(n) \rfloor} f_k.$$

For a fixed z, we want to show that the corresponding average converges. For $n \in \mathbb{N}$ large enough, we will approximate the sequence $\mathbf{1}_{A_z}(n)$ by sequences of the form $F(\{a_1(n)\})$, where F is a continuous function.

We establish this for $z = 0$ (the other cases follow similarly). Assume that $x(t)$ decreases to 0 (the other case) is similar, which means that $x(t)$ is eventually positive and also $\{x(t)\} = x(t)$ for t sufficiently large. In addition, we can also assume that c is positive. Observe that for $t \in A_0$, we have

$$\lfloor a_2(t) \rfloor = \lfloor a_1(t) \rfloor + \lfloor c \rfloor$$

by the definition of A_0. This is equivalent to the inequalities

$$\{a_1(t) + c\} + \{x(t)\} \leq 1$$
$$\{a_1(t)\} + \{c\} \leq 1$$

which can be condensed into

$$\{a_1(t)\} + \{x(t)\} \leq 1 - \{c\}, \tag{4.23}$$

since we assumed for simplicity that $x(t)$ is eventually positive. To summarize, we have shown that

$$n \in A_0 \iff \{a_1(n)\} + \{x(n)\} \leq 1 - \{c\} \tag{4.24}$$

Let $\varepsilon > 0$ be a small number. Since we have that the function $x(t)$ decreases to 0, we have that $\{x(t)\} < \varepsilon$ for t large enough. Consider the set

$$A_\varepsilon = \{n \in \mathbb{N}\colon \{a_1(n)\} \leq 1 - \{c\} - \varepsilon\}.$$

Then, for sufficiently large values of n, we observe that if $n \in A_\varepsilon$, then the inequality

$$\{a_1(t)\} + \{x(t)\} \leq 1 - \{c\}$$

holds as well. Namely, $A_\varepsilon \subseteq A_0$. Let us denote $B_\varepsilon = [0, 1 - c - \varepsilon]$ for convenience and observe that

$$\mathbf{1}_{A_\varepsilon}(n) = \mathbf{1}_{B_\varepsilon}(\{a_1(n)\}).$$

Now we approximate the function $\mathbf{1}_{B_\varepsilon}$ by a continuous function in the uniform norm, where $\mathbf{1}_{B_\varepsilon}$ is considered a function on the torus $\mathbb{T}$ in the natural way. We can define a continuous function on $\mathbb{T}$, such that F_ε agrees with $\mathbf{1}_{B_\varepsilon}$ on the set[7]

$$[\varepsilon, 1 - \{c\} - 2\varepsilon] \cup [1 - \{c\}, 1 - \varepsilon]$$

and such that $\|F_\varepsilon - \mathbf{1}_{B_\varepsilon}\|_\infty \leq 2$. We suppose that ε is small enough so that these intervals are well-defined. Observe that $\mathbf{1}_{B_\varepsilon}$ is equal to 1 on the first interval of this union and equal to 0 on the second interval.

[7] In the case that c is an integer, we make natural modifications to this set. For example, one could define the function F_ε so that it agrees with $\mathbf{1}_{\mathcal{B}_\varepsilon}$ on $[\varepsilon, 1 - 2\varepsilon]$. Basically, we only require the function F_ε to agree with $\mathbf{1}_{B_\varepsilon}$ on a set of measure $1 - O(\varepsilon)$ for our argument to work.

Observe that

$$A_0\setminus A_\varepsilon=\{n\in\mathbb{N}\colon 1-\{c\}-\varepsilon<\{a_1(n)\}\leq 1-\{c\}-\{x(n)\}\}\subseteq\{n\in\mathbb{N}\colon \{a_1(n)\}\in[1-\{c\}-\varepsilon,1-\{c\}]\}.$$

Since the function $a_1(t)$ is equidistributed modulo 1, we conclude that the set $A_0\setminus A_\varepsilon$ has upper density at most ε. Therefore, we have

$$\Big\|\mathop{\mathbb{E}}_{1\leq n\leq N}\mathbf{1}_{A_0}(n)\;T^{\lfloor a_1(n)\rfloor}(f_1\cdot T^{\lfloor c\rfloor}f_2)\cdot T^{\lfloor a_3(n)\rfloor}f_3\dots T^{\lfloor a_k(n)\rfloor}f_k-$$
$$\mathop{\mathbb{E}}_{1\leq n\leq N}\mathbf{1}_{B_\varepsilon}(\{a_1(n)\})\;T^{\lfloor a_1(n)\rfloor}(f_1\cdot T^{\lfloor c\rfloor}f_2)\cdot T^{\lfloor a_3(n)\rfloor}f_3\dots T^{\lfloor a_k(n)\rfloor}f_k\Big\|_{L^2(\mu)}\leq$$
$$\mathop{\mathbb{E}}_{1\leq n\leq N}\mathbf{1}_{A_0\setminus A_\varepsilon}(n)\,\big\|\mathbf{1}_{A_\varepsilon}-1_{A_\varepsilon\setminus A_0}\big\|_\infty\leq 2\varepsilon+o_N(1),\quad (4.25)$$

where we used the fact that $\mathbf{1}_{A_\varepsilon}(n)=\mathbf{1}_{B_\varepsilon}(\{a_1(n)\})$ for all $n\in\mathbb{N}$, the trivial bound for the values of $n\in A_0\setminus A_\varepsilon$ and the fact that the set $A_0\setminus A_\varepsilon$ has upper density at most ε.

We do a similar comparison for the averages weighted by $F_\varepsilon(\{a_1(n)\})$ and $1_{B_\varepsilon}(\{a_1(n)\})$. To be more specific, we reiterate that the functions 1_{B_ε} and F_ε agree on the set

$$[\varepsilon,1-\{c\}-2\varepsilon]\cup[1-\{c\},1-\varepsilon].$$

Accordingly, we have $\mathbf{1}_{B_\varepsilon}(\{a_1(n)\})=F_\varepsilon(\{a_1(n)\})$, unless

$$\{a_1(n)\}\in[0,\varepsilon)\cup(1-\{c\}-2\varepsilon,1-\{c\})\cup(1-\varepsilon,1).$$

Let C_ε denote the set of $n\in\mathbb{N}$ for which $\{a_1(n)\}$ belongs to this union. This union has measure 4ε, which implies that the upper density of C_ε is at most 4ε (since $a_1(n)$ is equidistributed modulo 1). Hence, we infer that

$$\Big\|\mathop{\mathbb{E}}_{1\leq n\leq N}F_\varepsilon(\{a_1(n)\})\;T^{\lfloor a_1(n)\rfloor}(f_1\cdot T^{\lfloor c\rfloor}f_2)\cdot T^{\lfloor a_3(n)\rfloor}f_3\dots T^{\lfloor a_k(n)\rfloor}f_k-$$
$$\mathop{\mathbb{E}}_{1\leq n\leq N}\mathbf{1}_{B_\varepsilon}(\{a_1(n)\})\;T^{\lfloor a_1(n)\rfloor}(f_1\cdot T^{\lfloor c\rfloor}f_2)\cdot T^{\lfloor a_3(n)\rfloor}f_3\dots T^{\lfloor a_k(n)\rfloor}f_k\Big\|_{L^2(\mu)}\leq$$
$$\mathop{\mathbb{E}}_{1\leq n\leq N}1_{C_\varepsilon}(n)\,\|F_\varepsilon-1_{B_\varepsilon}\|_\infty\leq 8\varepsilon+o_N(1),\quad (4.26)$$

where we utilized the fact that $\mathbf{1}_{B_\varepsilon}(\{a_1(n)\})=F_\varepsilon(\{a_1(n)\})$ for all n on the complement of C_ε, the trivial bound for the values of $n\in C_\varepsilon$ and the fact that C_ε has upper density at most 4ε.

Combining (4.25) and (4.26), we deduce that

$$\Big\|\mathop{\mathbb{E}}_{1\leq n\leq N}\mathbf{1}_{A_0}(n)\;T^{\lfloor a_1(n)\rfloor}(f_1\cdot T^{\lfloor c\rfloor}f_2)\cdot T^{\lfloor a_3(n)\rfloor}f_3\dots T^{\lfloor a_k(n)\rfloor}f_k-$$
$$\mathop{\mathbb{E}}_{1\leq n\leq N}F_\varepsilon(\{a_1(n)\})\;T^{\lfloor a_1(n)\rfloor}(f_1\cdot T^{\lfloor c\rfloor}f_2)\cdot T^{\lfloor a_3(n)\rfloor}f_3\dots T^{\lfloor a_k(n)\rfloor}f_k\Big\|_{L^2(\mu)}\leq 10\varepsilon+o_N(1).\quad (4.27)$$

Taking $\varepsilon\to 0$, we deduce that it is sufficient to verify that the averages

$$\mathop{\mathbb{E}}_{1\leq n\leq N}F(\{a_1(n)\})\;T^{\lfloor a_1(n)\rfloor}(f_1\cdot T^{\lfloor c\rfloor}f_2)\dots T^{\lfloor a_k(n)\rfloor}f_k$$

converge for any continuous function F on $\mathbb{T}$. This would imply that the averages

$$\mathop{\mathbb{E}}_{1\leq n\leq N}1_{A_0}(n)\;T^{\lfloor a_1(n)\rfloor}(f_1\cdot T^{\lfloor c\rfloor}f_2)\dots T^{\lfloor a_k(n)\rfloor}f_k$$

converge in norm.

After approximating F by trigonometric polynomials (in the uniform norm), it suffices to show that the averages

$$\mathop{\mathbb{E}}_{1\leq n\leq N}e(l_1a_1(n))\;T^{\lfloor a_1(n)\rfloor}(f_1\cdot T^{\lfloor c\rfloor}f_2)\dots T^{\lfloor a_k(n)\rfloor}f_k$$

converge in norm for any $l_1\in\mathbb{Z}$. Note that the function $a_2(t)$ has vanished and its role has been replaced by the sequence $e(l_1a_1(n))$.

We repeat this process until we eliminate all pairs of functions, whose difference grows slower than $\log t$, where at each step our averages are multiplied by a sequence of the form $e(l_i a_i(n))$ ($l_i \in \mathbb{Z}$). After finitely many iterations, our problem eventually reduces to the following: let $a_1, ..., a_k$ satisfy (1.10) or (1.11) and let $b_1, ..., b_m$ be a subset of $\{a_1, ..., a_k\}$, so that the functions $b_1, ..., b_m$ satisfy the hypotheses of Proposition 3.1.1. Then, for any integers $l_1, ...l_k$, the averages

$$\mathop{\mathbb{E}}_{1\leq n\leq N} e(l_1 a_1(n) + \cdots + l_k a_k(n))\; T^{\lfloor b_1(n)\rfloor} f_1 \ldots T^{\lfloor b_m(n)\rfloor} f_m$$

converge in $L^2(\mu)$ for all functions $f_1, \ldots, f_m \in L^\infty(\mu)$.

Now we can apply Proposition 3.1.1 and use a standard telescopic argument to show that the limiting behavior of the above averages does not change if we replace the functions f_i by their projections to the factor $Z_s(X)$ (the number s is the one given by Proposition 3.1.1). However, by Theorem G, the factors $Z_s(X)$ are inverse limits of s-step nilsystems. Thus, by another standard limiting argument, we may reduce to the case that the space X is a nilmanifold and μ is its Haar measure, while the transformation T is the action (by left multiplication) of an element g on X. Finally, we can approximate the functions f_i by continuous functions and reduce our problem to the following:

If $X = G/\Gamma$ is a nilmanifold with $g \in G$ and the functions $a_1, ..., a_k, b_1, ..., b_m \in \mathcal{H}$ are as above, then for any continuous functions $f_1, ..., f_m$ the averages

$$\mathop{\mathbb{E}}_{1\leq n\leq N} e(l_1 a_1(n) + \cdots + l_k a_k(n))\; f_1(g^{\lfloor b_1(n)\rfloor} x) \ldots f_k(g^{\lfloor b_m(n)\rfloor} x)$$

converge in mean.

We show that these averages converge pointwise for every $x \in X$. We recall that the functions $b_1, ..., b_m$ belong to the set $\{a_1, ..., a_k\}$ (this is the only thing that we will need to use for the rest of the proof).

First of all, it suffices to show that the averages

$$\mathop{\mathbb{E}}_{1\leq n\leq N} e(l_1 a_1(n) + \cdots + l_k a_k(n))\; f_1(g^{b_1(n)} x) \ldots f_k(g^{b_m(n)} x)$$

converge pointwise, where $X = G/\Gamma$ is such that G is connected, simply connected nilpotent Lie group (basically, we can remove the integer parts appearing in the iterates). This follows by standard modifications in the proof of Lemma 4.1.3 (the fact that we have the coefficients $e(l_1 a_1(n) + \cdots + l_k a_k(n))$ in the final expression does not affect the argument), so we omit the details.

Now, observe that we can write the above averages in the form

$$\mathop{\mathbb{E}}_{1\leq n\leq N} F_0(g_0^{l_1 a_1(n)+\cdots+l_k a_k(n)} \tilde{x})\; F_1(\tilde{g}^{b_1(n)} \tilde{x}) \ldots F_k(\tilde{g}^{b_m(n)} \tilde{x}),$$

where $g_0 = (1_{\mathbb{T}}, e_G)$ and $\tilde{g} = (1_{\mathbb{T}}, g)$ act on the product nilmanifold $\mathbb{T} \times X$, the point $\tilde{x}$ is just $(\mathbb{Z}, x)$ and the functions F_i are defined by

$$F_0(y\mathbb{Z}, a\Gamma) = e(y) \;\text{ and }\; F_i(y\mathbb{Z}, a\Gamma) = f_i(a\Gamma) \text{ for } i \geq 1.$$

These are continuous functions on $\mathbb{T} \times X$. The functions $l_1 a_1(t) + \cdots + l_k a_k(t), b_1(t), ..., b_m(t)$ satisfy the hypotheses of Theorem 1.2.2 (since the functions $a_1, ..., a_k$ do) and the result follows. □

Chapter 5

Proofs for results along primes

5.1 Details of the proof

Our main goal is to use the Taylor expansion to replace our functions with polynomial iterates and then apply Proposition 5.2.3 below. In order to do this, we need to eliminate the error term of the approximations, since Proposition 5.2.3 concerns polynomials only.

In the case of strongly non-polynomial functions that also grow faster than some fractional power, we show that the associated Taylor polynomial $p_N(n)$ has ideal equidistribution properties. Indeed, by picking the length $L(N)$ a little more carefully, one gains arbitrary logarithmic powers over the trivial bound in the exponential sums of p_N. Consequently, we get that the number of integers in $[N, N+L(N)]$ for which $\lfloor a(n) \rfloor \neq \lfloor p_N(n) \rfloor$ is less than $L(N)(\log N)^{-100}$ (say) and, thus, their contribution to the average is negligible. Therefore, for all intents and purposes, one can suppose that the error terms are identically zero.

The situation is different when a function that grows slower than all fractional powers is involved since these functions are practically constant in these short intervals. For instance, if one has the function $p(t) + \log^2 t$, where p is a polynomial, the only feasible approximation is of the form $p(n) + \log^2 n = p(n) + \log^2 N + e_N(n)$, where $e_N(n)$ converges to 0. While it seems that we do have a polynomial as the main term in the approximation (at least when p is non-constant), quantitative bounds on the exponential sums of the polynomial component cannot be established in this case at all. The main reason is that such bounds depend heavily on the diophantine properties of the coefficients of p, for which we have no data.

In the case that p is a constant polynomial, we can use the equidistribution (mod 1) of the sequence $\log^2 n$ to show that in most short intervals $[N, N+L(N)]$, we have $\lfloor \log^2 n \rfloor = \lfloor \log^2 N \rfloor$ for all $n \in [N, N+L(N)]$. The contribution of the bad short intervals is then bounded using the triangle inequality and Corollary 2.4.3.

Suppose that the polynomial p above is non-constant. In the case that p has rational non-constant coefficients, we split our averages to suitable arithmetic progressions so that the resulting polynomials have integer coefficients (aside from the constant term) and, thus, the effect of $e_N(n)$ will be eliminated when we calculate the integer parts. In the case that p has a non-constant irrational coefficient, we can invoke the well-distribution of $p(n)$ to conclude that the number of integers of the set

$$E_N = \{n \in [N, N+L(N)] \colon \lfloor p(n) + \log^2 n \rfloor \neq \lfloor p(n) + \log^2 N \rfloor\}$$

is $O(\varepsilon L(N))$, for a fixed small parameter ε and N large. However, in order to bound the total contribution of the set E_N, we can only use the triangle inequality in the corresponding ergodic averages, so we are forced to extract information on how large the quantity

$$\frac{1}{L(N)} \sum_{N \leq n \leq N+L(N)} \Lambda_{w,b}(n) \mathbf{1}_{E_N}(n)$$

can be. This can be bounded effectively if the corresponding exponential sums

$$\frac{1}{L(N)} \sum_{N \leq n \leq N+L(N)} \Lambda_{w,b}(n) e\big(p(n)\big)$$

are small. This is demonstrated by combining the fact that the exponential sums of $p(n)$ are small (due to the presence of an irrational coefficient) with the fact that exponential sums weighted by $\Lambda_{w,b}(n)-1$ are small due to the uniformity of the W-tricked von Mangoldt function. The conclusion follows again by an application of the Erdős-Turàn inequality, this time for a probability measure weighted by $\Lambda_{w,b}(n)$.

5.1.1 A model example

We sketch the main steps in the case of the ergodic averages

$$\frac{1}{N}\sum_{n=1}^{N}(\Lambda_{w,b}(n)-1)T^{\lfloor n\log n\rfloor}f_1\cdot T^{\left\lfloor an^2+\log n\right\rfloor}f_2\cdot T^{\left\lfloor \log^2 n\right\rfloor}f_3. \tag{5.1}$$

where a is an irrational number. We will show that the L^2-norm of this expression converges to 0, as $N\to+\infty$ and then $w\to+\infty$. Note that the three sequences in the iterates satisfy our hypotheses. In addition, we remark that the arguments below are valid in the setting where we have three commuting transformations, but we consider a simpler case for convenience. Additionally, we do not evaluate the sequences at $Wn+b$ (as we should in order to be in the setup of Theorem 1.3.1), since the underlying arguments remain identical apart from changes in notation.

We choose $L(t)=t^{0.66}$ (actually, any power t^c with $5/8<c<2/3$ works here) and claim that it suffices to show that

$$\mathop{\mathbb{E}}_{1\le r\le R}\Big\|\mathop{\mathbb{E}}_{r\le n\le r+L(r)}\big(\Lambda_{w,b}(n)-1\big)T^{\lfloor n\log n\rfloor}f_1\cdot T^{\left\lfloor an^2+\log n\right\rfloor}f_2\cdot T^{\left\lfloor \log^2 n\right\rfloor}f_3\Big\|_{L^2(\mu)}=0. \tag{5.2}$$

This reduction is the content of Lemma 5.4.1. Now, we can use the Taylor expansion around r to write

$$\begin{aligned}
n\log n &= r\log r+(\log r+1)(n-r)+\frac{(n-r)^2}{2r}-\frac{(n-r)^3}{6\xi_{1,n,r}^2}\\
\log n &= \log r+\frac{n-r}{\xi_{2,n,r}}\\
\log^2 n &= \log^2 r+\frac{2(n-r)\log\xi_{3,n,r}}{\xi_{3,n,r}},
\end{aligned}$$

for some real numbers $\xi_{i,n,r}\in[r,n]$ $(i=1,2,3)$. Our choice of $L(t)$ implies that

$$\Big|\frac{(n-r)^3}{6\xi_{1,n,r}^2}\Big|\le\frac{r^{3\cdot 0.65}}{6r^2}\ll 1,$$

and similarly for the other two cases. To be more specific, there exists a $\delta>0$, such that all the error terms (the ones involving the quantities $\xi_{i,n,r}$) are $O(r^{-\delta})$.

Let us fix a small $\varepsilon>0$. Firstly, we shall deal with the third iterate, since this is the simplest one. Observe that if r is chosen large enough and such that it satisfies $\{\log^2 r\}\in(\varepsilon,1-\varepsilon)$, then for all $n\in[r,r+L(r)]$, we will have

$$\left\lfloor\log^2 n\right\rfloor=\left\lfloor\log^2 r\right\rfloor,$$

since the error terms in the expansion are $O(r^{-\delta})$, which is smaller than ε for large r. In addition, the sequence $\log^2 n$ is equidistributed modulo 1, so our prior assumption can fail for at most $3\varepsilon R$ (say) values of $r\in[1,R]$, provided that R is sufficiently large. For the bad values of r, we use the triangle inequality for the corresponding norm to deduce that their contribution on the average is $O(\varepsilon R)$, which will be acceptable if ε is small. Actually, in order to establish this, we will need to use Corollary 2.4.3, though we will ignore that in this exposition. In conclusion, we can rewrite the expression in (5.2) as

$$\mathop{\mathbb{E}}_{1\le r\le R}\Big\|\mathop{\mathbb{E}}_{r\le n\le r+L(r)}\big(\Lambda_{w,b}(n)-1\big)T^{\lfloor n\log n\rfloor}f_1\cdot T^{\left\lfloor an^2+\log n\right\rfloor}f_2\cdot T^{\left\lfloor \log^2 r\right\rfloor}f_3\Big\|_{L^2(\mu)}+O(\varepsilon). \tag{5.3}$$

Now, we deal with the first function. We claim that the discrepancy of the finite sequence

$$\Big(\{r\log r+(\log r+1)(n-r)+\frac{(n-r)^2}{2r}\}\Big)_{r\leq n\leq r+L(r)}$$

is $O_A(\log^{-A} r)$ for any $A>0$. We will establish this in Proposition 5.3.1 using Lemma 2.3.7 and Theorem I. As a baby case, we show the following estimate for some simple trigonometric averages:

$$\Big|\mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} e\Big(\frac{(n-r)^2}{2r}\Big)\Big|\leq \frac{1}{\log^A r}$$

for r large enough. Indeed, if that inequality fails for some $r\in\mathbb{N}$, there exists an integer $|q_r|\leq \log^{O(A)} r$, such that

$$\Big\|\frac{q_r}{2r}\Big\|_{\mathbb{T}}\leq \frac{\log^{O(A)} r}{(L(r))^2}.$$

If r is large enough, we can replace the norm with the absolute value, so that the previous inequality implies that

$$\big(L(r)\big)^2\leq \frac{2r\log^{O(A)} r}{|q_r|}.$$

However, the choice $L(t)=t^{0.66}$ implies that this inequality is false for large r.

In our problem, we can just pick $A=2$. Using the definition of discrepancy, we deduce that the number of integers in $[r,r+L(r)]$, for which we have

$$\{r\log r+(\log r+1)(n-r)+\frac{(n-r)^2}{2r}\}\in[0,r^{-\delta/2}]\cup[1-r^{-\delta/2},1)$$

is $O(L(r)\log^{-2} r)$. However, if n does not belong to this set of bad values, we conclude that

$$\lfloor n\log n\rfloor=\Big\lfloor r\log r+(\log r+1)(n-r)+\frac{(n-r)^2}{2r}\Big\rfloor$$

since the error terms are $O(r^{-\delta})$. Furthermore, since $\Lambda_{w,b}(n)=O(\log r)$ for $n\in[r,r+L(r)]$, we conclude that the contribution of the bad values is $o_r(1)$ on the inner average. Therefore, we can rewrite the expression in (5.3) as

$$\mathop{\mathbb{E}}_{1\leq r\leq R}\Big\|\mathop{\mathbb{E}}_{r\leq n\leq r+L(r)}\big(\Lambda_{w,b}(n)-1\big)T^{\lfloor p_r(n)\rfloor}f_1\cdot T^{\lfloor an^2+\log n\rfloor}f_2\cdot T^{\lfloor \log^2 r\rfloor}f_3\Big\|_{L^2(\mu)}+O(\varepsilon)+o_R(1),\quad(5.4)$$

where $p_r(n)=r\log r+(\log r+1)(n-r)+\dfrac{(n-r)^2}{2r}$.

Finally, we deal with the second iterate. We consider the parameter ε as above and set $M=1/\varepsilon$. Once again, we shall assume that r is very large compared to M. Since a is irrational, we have that the sequence an^2 is well-distributed modulo 1, so we would expect the number of n for which $\{an^2+\log r\}\notin[\varepsilon,1-\varepsilon]$ to be small. Note that for the remaining values of n, we have $\lfloor an^2+\log n\rfloor=\lfloor an^2+\log r\rfloor$, since the error term in the approximation is $O(r^{-\delta})$. Therefore, we estimate the size of the set

$$\mathcal{B}_{r,\varepsilon}:=\{n\in[r,r+L(r)]\colon \{an^2+\log r\}\in[0,\varepsilon]\cup[1-\varepsilon,1)\}$$

Using Weyl's theorem, we conclude that

$$\max_{1\leq m\leq M}\Big|\mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} e\big(m(an^2+\log r)\big)\Big|=o_r(1).\quad(5.5)$$

Here, the $o_r(1)$ term depends on $M=1/\varepsilon$, but since we will send $r\to+\infty$ and then $\varepsilon\to 0$, this will not cause any issues. We suppress these dependencies in this exposition.

An application of Theorem I implies that

$$\frac{|\mathcal{B}_{r,\varepsilon}|}{L(r)}\ll 2\varepsilon+\frac{1}{M}+\sum_{m=1}^{M}\frac{1}{m}\Big|\mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} e\big(m(ar^2+\log r)\big)\Big|,\quad(5.6)$$

so that $|\mathcal{B}_{r,\varepsilon}| \ll (\varepsilon + o_r(1))L(r)$. Additionally, we will need to estimate

$$\frac{1}{L(r)} \sum_{r \le n \le r+L(r)} \Lambda_{w,b}(n) \mathbf{1}_{\mathcal{B}_r}(n),$$

which will arise when we apply the triangle inequality to bound the contribution of the set $\mathcal{B}_r$. However, we have that

$$\max_{1 \le m \le M} \Big| \mathop{\mathbb{E}}_{r \le n \le r+L(r)} \Lambda_{w,b}(n) e\big(m(an^2 + \log r)\big) \Big| = o_w(1) + o_r(1), \tag{5.7}$$

which can be seen by splitting $\Lambda_{w,b}(n) = (\Lambda_{w,b}(n) - 1) + 1$, applying the triangle inequality and using Lemma 2.4.2 and (5.5), respectively, to treat the resulting exponential averages. In view of this, we can apply the Erdős-Turán inequality (Theorem I) for the probability measure

$$\nu(S) = \frac{\sum_{r \le n \le r+L(r)} \Lambda_{w,b}(n) \delta_{\{an^2 + \log r\}}(S)}{\sum_{r \le n \le r+L(r)} \Lambda_{w,b}(n)}$$

as well as Corollary 2.4.3 (to bound the sum in the denominator) to conclude that

$$\frac{1}{L(r)} \sum_{r \le n \le r+L(r)} \Lambda_{w,b}(n) \mathbf{1}_{\mathcal{B}_r}(n) \ll \varepsilon + o_w(1) \log \frac{1}{\varepsilon} + o_r(1),$$

Therefore, if we apply the triangle inequality, we conclude that the contribution of the set $\mathcal{B}_{r,\varepsilon}$ on the average over $[r, r + L(r)]$ is at most $O(\varepsilon + o_w(1) \log \frac{1}{\varepsilon} + o_r(1))$. This is acceptable if we send $R \to +\infty$, then $w \to +\infty$, and then $\varepsilon \to 0$ at the end.

Ignoring the peculiar error terms, we can rewrite the expression in (5.4) as

$$\mathop{\mathbb{E}}_{1 \le r \le R} \Big\| \mathop{\mathbb{E}}_{r \le n \le r+L(r)} (\Lambda_{w,b}(n) - 1) T^{\lfloor p_r(n) \rfloor} f_1 \cdot T^{\lfloor an^2 + \log r \rfloor} f_2 \cdot T^{\lfloor \log^2 r \rfloor} f_3 \Big\|_{L^2(\mu)}. \tag{5.8}$$

Now, the iterates satisfy the assumptions of Proposition 5.2.3 below. This is true for the first iterate since we have a good bound on the discrepancy and it is also true for the second iterate because the polynomial an^2 has an irrational coefficient (so we can use its well-distribution modulo 1). For the third one, our claim is obvious because we simply have an integer in the iterate. Therefore, we can bound the inner average by a constant multiple of the norm

$$\|\Lambda_{w,b} - 1\|_{U^s(r, r+L(r)]}$$

with some error terms that we will ignore here. Finally, we invoke Theorem K to show that the average

$$\mathop{\mathbb{E}}_{1 \le r \le R} \|\Lambda_{w,b} - 1\|_{U^s(r, r+L(r)]}$$

converges to 0, which leads us to our desired conclusion.

5.2 Lifting to an extension flow

In this section, we use a trick that allows us to replace the polynomial ergodic averages with similar ergodic averages over $\mathbb{R}$ actions on an extension of the original probability space, removing the rounding functions in the process.. This argument is implicit in [33] for Cesàro averages, so we adapt its proof to the setting of short intervals. Firstly, we will need a Gowers norm bound for multiple ergodic averages with polynomial iterates. The following proposition is due to Frantzikinakis, Host, and Kra and follows from a PET induction scheme similar to the one we used in Section 3.4.5.

Proposition 5.2.1. *[20, Lemma 3.5] Let $k, \ell \in \mathbb{N}$, $(X, \mathcal{X}, \mu, T_1, \ldots, T_k)$ be a system of commuting $\mathbb{Z}$ actions, $p_{i,j} \in \mathbb{Z}[t]$ be polynomials for every $1 \leq i \leq k$, $1 \leq j \leq \ell$, $f_1, \ldots, f_\ell \in L^\infty(\mu)$ and $a : \mathbb{N} \to \mathbb{C}$ be a sequence. Then, there exists $s \in \mathbb{N}$, depending only on the maximum degree of the polynomials $p_{i,j}$ and the integers k, ℓ, and a constant C_s depending on s, such that*

$$\Big\| \mathop{\mathbb{E}}_{1 \leq n \leq N} a(n) \cdot \prod_{j=1}^{\ell} \prod_{i=1}^{k} T_i^{p_{i,j}(n)} f_j \Big\|_{L^2(\mu)} \leq C_s \left(\big\| a \cdot \mathbf{1}_{[1,N]} \big\|_{U^s(\mathbb{Z}_{sN})} + \frac{\max\{1, \|a\|^{2s}_{\ell^\infty[1,sN]}\}}{N} \right). \tag{5.9}$$

Remark. (i) The statement presented in [20] asserts that the second term in the prior sum is just $o_N(1)$, under the assumption that $a(n) \ll n^c$ for all $c > 0$. However, a simple inspection of the proof gives the error term presented above. Indeed, the error terms appearing in the proof of Proposition 5.2.1 are precisely of the form

$$\frac{1}{N} \mathop{\mathbb{E}}_{n \in [1,N]} \mathop{\mathbb{E}}_{\underline{h} \in [1,N]^k} \Big| \prod_{\underline{\varepsilon} \in \{0,1\}^k} \mathcal{C}^{|\underline{\varepsilon}|} \, a(n + \underline{h} \cdot \underline{\varepsilon}) \Big|$$

for $k \leq s-1$, which are the error terms in the van der Corput inequality. Deducing the error term on (5.9) is then straightforward.
(ii) The number $s-1$ is equal to the number of applications of the van der Corput inequality in the associated PET argument and we may always assume that $s \geq 2$. In that case, Lemma 2.2.1 and the bound $\big\|\mathbf{1}_{[1,N]}\big\|_{U^s(\mathbb{Z}_{sN})} \leq 1$ implies that we can replace the norm in (5.9) with the term $\|a\|_{U^s[1,N]}$.

For polynomials $p_{i,j}(t) \in \mathbb{R}[t]$ of the form

$$p_{i,j}(t) = a_{ij,d_{ij}} t^{d_{ij}} + \cdots + a_{ij,1} t + a_{ij,0},$$

and $(T_{i,s})_{s \in \mathbb{R}}$ $\mathbb{R}$-actions, we have

$$T_{i, p_{i,j}(n)} = \left(T_{i, a_{ij,d_{ij}}}\right)^{n^{d_{ij}}} \cdot \ldots \cdot \left(T_{i, a_{ij,1}}\right)^{n} \cdot \left(T_{i, a_{ij,0}}\right).$$

Thus, Proposition 5.2.1 implies the following.

Corollary 5.2.2. *Let $k, \ell \in \mathbb{N}$, $(X, \mathcal{X}, \mu, S_1, \ldots, S_k)$ be a system of commuting $\mathbb{R}$-actions, $p_{i,j} \in \mathbb{Z}[t]$ be polynomials for all $1 \leq i \leq k$, $1 \leq j \leq \ell$, $f_1, \ldots, f_\ell \in L^\infty(\mu)$ and $a : \mathbb{N} \to \mathbb{C}$ be a sequence. Then, there exists $s \in \mathbb{N}$, depending only on the maximum degree of the polynomials $p_{i,j}$ and the integers k, ℓ and a constant C_s depending on s, such that*

$$\Big\| \mathop{\mathbb{E}}_{1 \leq n \leq N} a(n) \cdot \prod_{j=1}^{\ell} \prod_{i=1}^{k} S_{i, p_{i,j}(n)} f_j \Big\|_{L^2(\mu)} \leq C_s \left(\big\| a \cdot \mathbf{1}_{[1,N]} \big\|_{U^s(\mathbb{Z}_{sN})} + \frac{\max\{1, \|a\|^{2s}_{\ell^\infty[1,sN]}\}}{N} \right). \tag{5.10}$$

The following proposition allows us to obtain Gowers norm bounds for averages with real polynomial iterates by using a trick to replace our $\mathbb{Z}^d$-action with a $\mathbb{R}^d$-action.

Proposition 5.2.3. *Let k, ℓ, d be positive integers and let $L(N)$ be a positive sequence satisfying $N^{\frac{5}{8}+\varepsilon} \ll L(N) \ll N^{1-\varepsilon}$. Let $(X, \mathcal{X}, \mu, T_1, \ldots, T_k)$ be a system of commuting transformations. Then, there exists a positive integer s depending only on k, ℓ, d, such that for any variable family $\mathcal{P} = \{p_{i,j,N} \colon 1 \leq i \leq k, 1 \leq j \leq \ell\}$ of polynomials with degrees at most d that, for all i, j, satisfy*

$$\lim_{\delta \to 0^+} \lim_{N \to +\infty} \frac{|\{N \leq n \leq N + L(N) : \; \{p_{i,j,N}(n)\} \in [1-\delta, 1)\}|}{L(N)} = 0, \tag{5.11}$$

we have that for any $0 < \delta < 1$ and functions $f_1, \ldots, f_\ell \in L^\infty(\mu)$

$$\Big\| \mathop{\mathbb{E}}_{N \leq n \leq N+L(N)} \left(\Lambda_{w,b}(n) - 1\right) \prod_{j=1}^{\ell} \prod_{i=1}^{k} T_i^{\lfloor p_{i,j,N}(n) \rfloor} f_j \Big\|_{L^2(\mu)} \ll_{k,\ell,d}$$
$$\frac{1}{\delta^{k\ell}} \Big(\big\| \Lambda_{w,b}(n) - 1 \big\|_{U^s(N, N+sL(N)]} + o_w(1) \Big) + o_\delta(1)(1 + o_w(1)),$$

for all $1 \leq b \leq W$, $(b, W) = 1$, where $W = \prod_{p \in \mathbb{P}\colon p \leq w} p$.

Proof. Let λ denote the Lebesgue measure on $[0,1)$ and we define (as in [33]) the measure-preserving $\mathbb{R}^{k\ell}$-action $\prod_{i=1}^{k} S_{i,s_{i,1}} \cdot \ldots \cdot \prod_{i=1}^{k} S_{i,s_{i,\ell}}$ on the space $Y := X \times [0,1)^{k\ell}$, endowed with the measure $\nu := \mu \times \lambda^{k\ell}$, by

$$\prod_{j=1}^{\ell}\prod_{i=1}^{k} S_{i,s_{i,j}}(x, a_{1,1}, \ldots, a_{k,1}, a_{1,2}, \ldots, a_{k,2}, \ldots, a_{1,\ell}, \ldots, a_{k,\ell}) =$$

$$\left(\prod_{j=1}^{\ell}\prod_{i=1}^{k} T_i^{[s_{i,j}+a_{i,j}]} x, \{s_{1,1}+a_{1,1}\}, \ldots, \{s_{k,1}+a_{k,1}\}, \ldots, \{s_{1,\ell}+a_{1,\ell}\}, \ldots, \{s_{k,\ell}+a_{k,\ell}\} \right).$$

If $f_1, \ldots, f_\ell$ are bounded functions on X, we define the Y-extensions of f_j, setting for every element $(a_{1,1}, \ldots, a_{k,1}, a_{1,2}, \ldots, a_{k,2}, \ldots, a_{1,\ell}, \ldots, a_{k,\ell}) \in [0,1)^{k\ell}$:

$$\hat{f}_j(x, a_{1,1}, \ldots, a_{k,1}, a_{1,2}, \ldots, a_{k,2}, \ldots, a_{1,\ell}, \ldots, a_{k,\ell}) = f_j(x), \;\; 1 \le j \le \ell;$$

and we also define the function

$$\hat{f}_0(x, a_{1,1}, \ldots, a_{k,1}, a_{1,2}, \ldots, a_{k,\ell}) = 1_{[0,\delta]^{k\ell}}(a_{1,1}, \ldots, a_{k,1}, a_{1,2}, \ldots, a_{k,\ell}).$$

For every $N \le n \le N + L(N)$, we consider the functions (on the original space X)

$$b_N(n) := (\prod_{i=1}^{k} T_i^{[p_{i,1,N}(n)]}) f_1 \cdot \ldots \cdot (\prod_{i=1}^{k} T_i^{[p_{i,\ell,N}(n)]}) f_\ell$$

as well as the functions

$$\tilde{b}_N(n) := \hat{f}_0 \cdot (\prod_{j=1}^{\ell}\prod_{i=1}^{k} S_{i,\delta_{j1}\cdot p_{i,1,N}(n)}) \hat{f}_1 \cdot \ldots \cdot (\prod_{j=1}^{\ell}\prod_{i=1}^{k} S_{i,\delta_{j\ell}\cdot p_{i,\ell,N}(n)}) \hat{f}_\ell$$

defined on the extension Y. Here, δ_{ij} denotes the Kronecker δ, meaning that the only terms that do not vanish are the diagonal ones (i.e., when $i = j$). For every $x \in X$, we also let

$$b'_N(n)(x) := \int_{[0,1)^{k\ell}} \tilde{b}_N(n)(x, a_{1,1}, \ldots, a_{k,1}, a_{1,2}, \ldots, a_{k,2}, \ldots, a_{1,\ell}, \ldots, a_{k,\ell}) \, d\lambda^{k\ell},$$

where the integration is with respect to the variables $a_{i,j}$.

Using the triangle and Cauchy-Schwarz inequalities, we have

$$\delta^{k\ell} \Big\| \mathbb{E}_{N \le n \le N+L(N)} \big(\Lambda_{w,b}(n) - 1\big) b_N(n) \Big\|_{L^2(\mu)} \le$$
$$\Big\| \mathbb{E}_{N \le n \le N+L(N)} \big(\Lambda_{w,b}(n) - 1\big) \cdot (\delta^{k\ell} b_N(n) - b'_N(n)) \Big\|_{L^2(\mu)} + \Big\| \mathbb{E}_{N \le n \le N+L(N)} \big(\Lambda_{w,b}(n) - 1\big) \tilde{b}_N(n) \Big\|_{L^2(\nu)}. \tag{5.12}$$

Using Proposition 5.2.1, we find an integer $s \in \mathbb{N}$, depending only on the integers k, ℓ, d, and a constant C_s depending on s, such that

$$\Big\| \mathbb{E}_{N \le n \le N+L(N)} \big(\Lambda_{w,b}(n) - 1\big) \tilde{b}_N(n) \Big\|_{L^2(\nu)} \le C_s \left(\|\Lambda_{w,b} - 1\|_{U^s(N, N+sL(N)]} + o_N(1) \right), \tag{5.13}$$

where the $o_N(1)$ term depends only on the integer s and the sequence $\Lambda_{w,b}(n)$.

Now we study the first term

$$\Big\| \mathbb{E}_{N \le n \le N+L(N)} \big(\Lambda_{w,b}(n) - 1\big) \cdot (\delta^{k\ell} b_N(n) - b'_N(n)) \Big\|_{L^2(\mu)}$$

in (5.12). For every $x \in X$ and $N \le n \le N + L(N)$, we have

$$\left|\delta^{k\ell} b_N(n)(x) - b'_N(n)(x)\right| = \left|\int_{[0,\delta]^{k\ell}} \left(\prod_{j=1}^{\ell} f_j(\prod_{i=1}^{k} T_i^{[p_{i,j,N}(n)]}x) - \prod_{j=1}^{\ell} f_j(\prod_{i=1}^{k} T_i^{[p_{i,j,N}(n)+a_{i,j}]}x)\right) d\lambda^{k\ell}\right|.$$

Since all the integrands $a_{i,j}$ are less than or equal than δ, we deduce that if all of the implicit polynomials satisfy $\{p_{i,j,N}(n)\} < 1-\delta$, we have $T_i^{[p_{i,j,N}(n)+a_{i,j}]} = T_i^{[p_{i,j,N}(n)]}$ for all $1 \le i \le k$, $1 \le j \le \ell$. To deal with the possible case where $\{p_{i,j,N}(n)\} \ge 1-\delta$ for at least one of our polynomials, we define, for every $1 \le i \le k$, $1 \le j \le \ell$, the set

$$E^{i,j}_{\delta,N} := \{n \in [N, N + L(N)] \colon \{p_{i,j,N}(n)\} \in [1-\delta, 1)\}.$$

Then, by using the fact that

$$\mathbf{1}_{E^{1,1}_{\delta,N} \cup \ldots \cup E^{1,\ell}_{\delta,N} \cup E^{2,1}_{\delta,N} \cup \ldots \cup E^{k,\ell}_{\delta,N}} \le \sum_{(i,j)\in[1,k]\times[1,\ell]} \mathbf{1}_{E^{i,j}_{\delta,N}}$$

and that $\mathbf{1}_{E^{i,j}_{\delta,N}}(n) = \mathbf{1}_{[1-\delta,1)}(\{p_{i,j,N}(n)\})$, we infer that

$$\left|\delta^{k\ell} b_N(n)(x) - b'_N(n)(x)\right| \le 2\delta^{k\ell} \sum_{(i,j)\in[1,k]\times[1,\ell]} \mathbf{1}_{[1-\delta,1)}(\{p_{i,j,N}(n)\})$$

for every $x \in X$. In view of the above, using the inequality $|\Lambda_{w,b}(n) - 1| \le \Lambda_{w,b}(n) + 1$, we deduce that

$$\begin{aligned} &\mathop{\mathbb{E}}_{N\le n\le N+L(N)} \left|\big(\Lambda_{w,b}(n) - 1\big)\right| \cdot \mathbf{1}_{[1-\delta,1)}(\{p_{i,j,N}(n)\}) \le \\ &\quad \mathop{\mathbb{E}}_{N\le n\le N+L(N)} \big(\Lambda_{w,b}(n) - 1\big) \cdot \mathbf{1}_{[1-\delta,1)}(\{p_{i,j,N}(n)\}) + 2 \mathop{\mathbb{E}}_{N\le n\le N+L(N)} \mathbf{1}_{[1-\delta,1)}(\{p_{i,j,N}(n)\}) \le \\ &\qquad \mathop{\mathbb{E}}_{N\le n\le N+L(N)} \big(\Lambda_{w,b}(n) - 1\big) \cdot \mathbf{1}_{[1-\delta,1)}(\{p_{i,j,N}(n)\}) + 2 \cdot \frac{|E^{i,j}_{\delta,N}|}{L(N)}. \end{aligned}$$

Since each polynomial $p_{i,j,N}$ satisfies (5.11) for large N and small enough δ, the term (and the sum of finitely many terms of this form) $\frac{|E^{i,j}_{\delta,N}|}{L(N)}$ is as small as we want.

It remains to show that the term

$$\mathop{\mathbb{E}}_{N\le n\le N+L(N)} \big(\Lambda_{w,b}(n) - 1\big) \cdot \mathbf{1}_{[1-\delta,1)}(\{p_{i,j,N}(n)\})$$

goes to zero as $N \to \infty$, then $w \to \infty$ and finally $\delta \to 0^+$. To this end, it suffices to show

$$\mathop{\mathbb{E}}_{N\le n\le N+L(N)} \big(\Lambda_{w,b}(n) - 1\big) e^{2\pi i m p_{i,j,N}(n)} \to 0$$

as $N \to \infty$ and then $w \to \infty$ for all $m \in \mathbb{Z} \setminus \{0\}$,[1] which follows from Lemma 2.4.2. □

5.3 Equidistribution in short intervals

We gather here some useful propositions that describe the behavior of a Hardy field function when restricted to intervals of the form $[N, N+L(N)]$, where $L(N)$ grows slower compared to the parameter N. In our applications, we will typically need the function $L(N)$ to grow faster than $N^{5/8}$ in order to be able to use the uniformity results in short intervals, but we will not need to work under this assumption throughout most of this section, the only exception being Proposition 5.3.3 below. We will also present an example that illustrates the main points in the proof of Theorem 1.3.1 in the following section.

[1]This follows by the fact that if f is Riemann integrable on $[0, 1)$ with $\int_{[0,1)} f(x)\,dx = c$, then, for every $\varepsilon > 0$, we can find trigonometric polynomials q_1, q_2, with no constant terms, with $q_1(t) + c - \varepsilon \le f(t) \le q_2(t) + c + \varepsilon$. We use this for the function $f = \mathbf{1}_{[1-\delta,1)}$.

5.3.1 Eliminating the error terms in the approximations

In the previous chapters, we saw that any Hardy field function can be approximated by polynomials in short intervals using the Taylor expansion. Namely, if $a(t)$ diverges and $L(t) \to +\infty$ is a positive function, such that

$$\left|a^{(k)}(t)\right|^{-\frac{1}{k}} \prec L(t) \prec \left|a^{(k+1)}(t)\right|^{-\frac{1}{k+1}} \tag{5.14}$$

then, for any $0 \leq h \leq L(N)$, we have

$$a(N+h) = a(N+h) = a(N) + \cdots + \frac{h^k a^{(k)}(N)}{k!} + \frac{h^{k+1} a^{(k+1)}(\xi_{N,h})}{(k+1)!} = p_N(h) + \theta_N(h)$$

for some $\xi_{N,h} \in [N, N+h]$, where we denote

$$p_N(h) = a(N) + \cdots + \frac{h^k a^{(k)}(N)}{k!}.$$

Observe that our growth assumption on $L(t)$ implies that the term $\theta_N(h)$ is bounded by a quantity that converges to 0, as $N \to +\infty$. Therefore, for large values of N, we easily deduce that

$$\lfloor a(N+h) \rfloor = \lfloor p_N(h) \rfloor + \varepsilon_{N,h},$$

where $\varepsilon_{N,h} \in \{-1, 0, 1\}$. In order to be able to apply Proposition 5.2.3, we will need to eliminate the error terms $\varepsilon_{N,h}$. We will consider three distinct cases, which are tackled using somewhat different arguments.

The case of fast-growing functions

Firstly, we establish the main proposition that will allow us to remove the error terms in the case of functions that contain a "non-polynomial part" which does not grow too slowly. We will need a slight strengthening of the growth conditions in (5.14), which, as we saw previously, are sufficient to have a Taylor approximation in the interval $[N, N+L(N)]$.

Proposition 5.3.1. *Let $A > 0$ and let $a(t)$ be a C^∞ function defined for all sufficiently large $t \in \mathbb{R}$. Assume $L(t)$ is a positive sub-linear function going to infinity and let k be a positive integer, such that*

$$1 \lll \left|a^{(k)}(t)\right|^{-\frac{1}{k}} \lll L(t) \lll \left|a^{(k+1)}(t)\right|^{-\frac{1}{k+1}} \tag{5.15}$$

and such that the function $a^{(k+1)}(t)$ converges to 0 monotonically. Then, for N large enough, we have that, for all $0 \leq c \leq d < 1$,

$$\frac{\left|\{n \in [N, N+L(N)] \colon a(n) \in [c,d]\}\right|}{L(N)} = |d-c| + O_A(L(N)\log^{-A} N).^{2} \tag{5.16}$$

Consequently, for all N sufficiently large, we have that

$$\lfloor a(N+h) \rfloor = \left\lfloor a(N) + ha'(N) + \cdots + \frac{h^k a^{(k)}(N)}{k!} \right\rfloor$$

for all, except at most $O_A(L(N)\log^{-A}(N))$ values of integers $h \in [N, N+L(N)]$.

Proof. Our hypothesis on $L(t)$ implies that there exist $\varepsilon_1, \varepsilon_2 > 0$ such that

$$L(t)\left|a^{(k)}(t)\right|^{\frac{1}{k}} \gg t^{\varepsilon_1} \text{ and } L(t)\left|a^{(k+1)}(t)\right|^{\frac{1}{k+1}} \ll t^{-\varepsilon_2}. \tag{5.17}$$

[2] One can actually get a small power saving here, with an exponent that depends on k and the implicit fractional powers in the growth relations of (5.15), though this will not be any more useful for our purposes.

In addition, the leftmost inequality implies that there exists $\varepsilon_3 > 0$, such that $a^{(k)}(t) \ll t^{-\varepsilon_3}$. Using the Taylor expansion around the point N, we can write

$$a(N+h) = a(N) + ha'(N) + \cdots + \frac{h^k a^{(k)}(N)}{k!} + \frac{h^{k+1} a^{(k+1)}(\xi_h)}{(k+1)!}, \quad \text{for some } \xi_h \in [N, N+h], \quad (5.18)$$

for every $h \in [0, L(N)]$. We denote

$$p_N(h) = a(N) + \cdots + \frac{h^k a^{(k)}(N)}{k!}$$

and

$$\theta_N(h) = \frac{h^{k+1} a^{(k+1)}(\xi_h)}{(k+1)!}.$$

The function $a^{(k+1)}(t)$ converges to 0 monotonically due to our hypothesis. Therefore, for sufficiently large N,

$$\max_{0 \le h \le L(N)} |\theta_N(h)| \le \left| \frac{a^{(k+1)}(N)}{(k+1)!} \right| (L(N))^{k+1} = \theta_N, \quad (5.19)$$

and the quantity θ_N is strongly dominated by the constant 1 due to (5.17). More precisely, we have that $\theta_N \ll N^{-(k+1)\varepsilon_2}$.

Let $A > 0$ be any constant. We study the discrepancy of the finite polynomial sequence $(p_N(h))_{0 \le h \le L(N)}$. We shall establish that we have

$$\Delta_{[c,d]}\big(p_N(h)\big) \ll_A \log^{-A} N$$

for any choice of the interval $[c,d] \subseteq [0,1]$. To this end, we apply Theorem I for the finite sequence $(p_N(h))_{0 \le h \le L(N)}$ to deduce that

$$\Delta_{[c,d]}\Big(\big(p_N(h)\big)_{0 \le h \le L(N)}\Big) \le \frac{C}{\lfloor \log^A N \rfloor} + C \sum_{m=1}^{\lfloor \log^A N \rfloor} \frac{1}{m} \Big| \mathop{\mathbb{E}}_{0 \le h \le L(N)} e(m p_N(h)) \Big|, \quad (5.20)$$

where C is an absolute constant. We claim that for every $1 \le m \le \lfloor \log^A N \rfloor$, we have that

$$\Big| \mathop{\mathbb{E}}_{0 \le h \le L(N)} e(m p_N(h)) \Big| \le \frac{1}{\log^A N}, \quad (5.21)$$

provided that N is sufficiently large. Indeed, assume for the sake of contradiction that there exists $1 \le m_0 \le \lfloor \log^A N \rfloor$, such that

$$\Big| \mathop{\mathbb{E}}_{0 \le h \le L(N)} e(m_0 p_N(h)) \Big| > \frac{1}{\log^A N}. \quad (5.22)$$

The leading coefficient of $m_0 p_N(h)$ is equal to

$$\frac{m_0 a^{(k)}(N)}{k!}.$$

Then, Lemma 2.3.7 implies that there exists a constant C_k (depending only on k) an integer q satisfying $|q| \le \log^{C_k A} N$ and such that

$$\Big\| q \cdot \frac{m_0 a^{(k)}(N)}{k!} \Big\|_{\mathbb{T}} \le \frac{\log^{C_k A} N}{\lfloor L(N) \rfloor^k}.$$

The number $q m_0$ is bounded in magnitude by $\log^{(C_k+1)A}(N)$, so that

$$q \cdot \frac{m_0 a^{(k)}(N)}{k!} \ll \log^{(C_k+1)A} N \cdot N^{-\varepsilon_3} = o_N(1).$$

Therefore, for large values of N, we can substitute the circle norm of the fraction in (5.22) with the absolute value, which readily implies that

$$\left|q\cdot\frac{m_0a^{(k)}(N)}{k!}\right|\le\frac{\log^{C_kA}N}{\lfloor L(N)\rfloor^k}\implies \lfloor L(N)\rfloor^k\big|a^{(k)}(N)\big|\le k!\log^{C_kA}N.$$

However, this implies that $L(t)$ cannot strongly dominate the function $\left(a^{(k)}(t)\right)^{-\frac{1}{k}}$, which is a contradiction due to our hypothesis.

We have established that for every $1\le m\le\lfloor\log^A N\rfloor$ and large N, inequality (5.21) holds. Substituting this in (5.20), we deduce that

$$\Delta_{[c,d]}\Big(\big(p_N(h)\big)_{0\le h\le L(N)}\Big)\le\frac{C}{\lfloor\log^A N\rfloor}+C\sum_{m=1}^{\lfloor\log^A N\rfloor}\frac{1}{m\log^A N},$$

which implies that

$$\Delta_{[c,d]}\Big(\big(p_N(h)\big)_{0\le h\le L(N)}\Big)\ll\frac{A\log\log N}{\log^A N}.$$

In particular, since A was arbitrary, we get

$$\Delta_{[c,d]}\Big(\big(p_N(h)\big)_{0\le h\le L(N)}\Big)\ll_A\frac{1}{\log^A N}. \tag{5.23}$$

This establishes the first part of the proposition.

The second part of our statement follows from an application of the bound on the discrepancy of the finite polynomial sequence $(p_N(h))$. Indeed, we consider the set

$$S_N=[0,\theta_N]\cup[1-\theta_N,1),$$

where we recall that θ_N was defined in (5.19) and decays faster than a small fractional power. Then, if $\{p_N(h)\}\notin S_N$, we have $\lfloor p_N(h)+\theta_N(h)\rfloor=\lfloor p_N(h)\rfloor$, as can be seen by noticing that the error term in (5.18) is bounded in magnitude by θ_N. Now, we estimate the number of integers $h\in[0,L(N)]$ for which $\{p_N(h)\}\in S_N$.

Using the definition of discrepancy and the recently established bounds, we deduce that

$$\frac{\big|\{h\in[0,L(N)]\colon\{p_N(h)\}\in[0,\theta_N]\}\big|}{L(N)}-\theta_N\ll_A\frac{1}{\log^A N}$$

for every $A>0$. Since the number θ_N is dominated by $N^{-(k+1)\varepsilon_2}$, this implies that

$$\big|\{h\in[0,L(N)]\colon\{p_N(h)\}\in[0,\theta_N]\}\big|\ll_A\frac{L(N)}{\log^A N}.$$

An entirely similar argument yields the analogous relation for the interval $[1-\theta_N,1)$. Therefore, the number of integers in $[0,L(N)]$ for which $\{p_N(h)\}\in S_N$ is at most $O_A(L(N)\log^{-A}N)$.

In conclusion, since $\lfloor a(N+h)\rfloor=\lfloor p_N(h)\rfloor$ for all integers not in S_N, we have that the number of integers which does not satisfy this last relation is $O_A(L(N)\log^{-A}N)$, which yields the desired result. □

The above proposition asserts that, for almost all values of $h\in[0,L(N)]$, we can write $\lfloor a(N+h)\rfloor=\lfloor p_N(h)\rfloor$. The logarithmic power saving in the statement will be helpful since we are dealing with averages weighted by the sequence $\Lambda_{w,b}(n)-1$, which has size comparable to $\log N$ on the interval $[N,N+L(N)]$. Furthermore, notice that we did not assume that a is a Hardy field function in the proof. Thus, the conditions in this proposition can be used to prove a comparison result for more general iterates.

The case of slow functions

Unfortunately, the previous proposition cannot deal with functions whose only possible Taylor approximations involve only a constant term. This case will emerge when we have sub-fractional functions since, as we have already remarked, these functions have a polynomial approximation of degree 0 in short intervals (assuming that $L(t) \ll t^{1-\varepsilon}$). To cover this case, we will need the following proposition which is practically of a qualitative nature.

Proposition 5.3.2. *Let $a(t) \in \mathcal{H}$ be a sub-fractional function such that $a(t) \succ \log t$. Assume $L(t)$ is a positive sub-linear function going to infinity and such that $L(t) \ll t^{1-\delta}$, for some $\delta > 0$. Then, for every $0 < \varepsilon < 1$, we have the following: for all $R \in \mathbb{N}$ sufficiently large we have $\lfloor a(N+h) \rfloor = \lfloor a(N) \rfloor$ for every $h \in [0, L(N)]$, for all, except at most εR values of $N \in [1, R]$.*

Proof. Observe that for any $h \in [0, L(N)]$, we have

$$a(N+h) = a(N) + ha'(\xi_h) \tag{5.24}$$

for some $\xi_h \in [N, N+h]$. In addition, since $a'(t)$ converges to 0 monotonically, we have

$$|ha'(\xi_h)| \leq L(N)a'(N) \ll N^{1-\delta}a'(N) \lll 1,$$

where the last inequality follows from Lemma 2.1.3 and the assumption that $a(t)$ is sub-fractional. In particular, there exists a positive real number q, such that $|ha'(\xi_h)| \ll N^{-q}$, for all $h \in [0, L(N)]$.[3]

The sequence $a(n)$ is equidistributed mod 1 by Theorem E, since it dominates the function $\log t$. Now, suppose that $\varepsilon > 0$, and choose a number R_0 such that $R_0^{-2q} < \varepsilon/2$. Then, for $R \geq R_0$, the number of integers $N \in [R_0, R]$ such that $\{a(N)\} \in [\frac{\varepsilon}{2}, 1 - \frac{\varepsilon}{2}]$ is

$$(R - R_0)(1 - \varepsilon + o_R(1))$$

due to the fact that $a(n)$ is equidistributed. For these values of N, we have that

$$\{a(N)\} \notin [0, N^{-2q}] \cup [1 - N^{-2q}, 1],$$

which implies that for all $h \in [0, L(N)]$, we have that $\lfloor a(N+h) \rfloor = \lfloor a(N) \rfloor$, as can be derived easily by (5.24) and the fact that the error term is $O(N^{-q})$. If we consider the integers N in the interval $[1, R_0]$ as well, then the number of "bad values" (that is, the numbers N for which we do not have $\lfloor a(N+h) \rfloor = \lfloor a(N) \rfloor$ for every $h \in [0, L(N)]$) is at most

$$R_0 + (R - R_0)(\varepsilon + o_R(1)).$$

Finally, choosing R sufficiently large, we get that this number is smaller than $2\varepsilon R$ and the claim follows. □

In simplistic terms, what we have established is that if we restrict our attention to short intervals $[N, N + L(N)]$ for the natural numbers N, such that $\{a(N)\} \in [\varepsilon, 1 - \varepsilon]$, then we can just write $\lfloor a(N+h) \rfloor = \lfloor a(N) \rfloor$ for all $h \in [0, L(N)]$. Due to the equidistribution of $a(n)$ mod 1 (which follows from Theorem E), this is practically true for almost all N, if we take ε sufficiently small.

The case of polynomial functions

The final case is the case of functions of the form $p(t) + x(t)$, where p is a polynomial with real coefficients and $x(t)$ is a sub-fractional function. The equidistribution of the corresponding sequence will be affected only by the polynomial p when restricted to short intervals. Nonetheless, the techniques of Proposition 5.3.1 cannot be employed, because we cannot establish quantitative bounds on the exponential sums uniformly over all real polynomials. Therefore, we will use the following proposition, which allows us to calculate the integer parts in this case. Unlike the previous two propositions which can be bootstrapped to give a similar statement for several functions, we establish this one for several functions from the outset. We do not need to concern ourselves with rational polynomials, since these can be trivially reduced to the case of integer polynomials by passing to arithmetic progressions.

[3] We do not actually need this quantity to converge to zero faster than some power of N. The same argument applies if this quantity simply converges to zero.

Proposition 5.3.3. *Let k, d be positive integers, let $0 < \varepsilon < 1/2$ be a real number and let $w \in \mathbb{N}$. We define $W = \prod_{p\in\mathbb{P}:\, p\leq w} p$ and let $1 \leq b \leq W$ be any integer with $(b, W) = 1$. Suppose that $a_1, \dots, a_k \in \mathcal{H}$ are functions of the form $p_i(t) + x_i(t)$, where p_i are polynomials of degree at most d and with at least one irrational non-constant coefficient, while $x_i(t)$ are sub-fractional functions. Finally, assume that $L(t)$ is a positive sub-linear function going to infinity and such that*

$$t^{\frac{5}{8}} \lll L(t) \lll t.$$[4]

Then, for every r sufficiently large in terms of w, $\frac{1}{\varepsilon}$, we have that there exists a subset $\mathcal{B}_{r,\varepsilon}$ of integers in the interval $[r, r + L(r)]$ with at most $O_k(\varepsilon L(r))$ elements, such that for all integers $n \in [r, r + L(r)] \setminus \mathcal{B}_{r,\varepsilon}$, we have

$$\lfloor p_i(n) + x_i(n) \rfloor = \lfloor p_i(n) + x_i(r) \rfloor.$$

Furthermore, the set $\mathcal{B}_{r,\varepsilon}$ satisfies

$$\frac{1}{L(r)} \sum_{r\leq n\leq r+L(r)} \Lambda_{w,b}(n)\mathbf{1}_{\mathcal{B}_{r,\varepsilon}}(n) \ll_{k,d} \varepsilon + o_w(1)\log\frac{1}{\varepsilon} + o_r(1). \tag{5.25}$$

Remark. The $o_r(1)$ term depends on the fixed parameters w, ε. However, in our applications, we will send $r \to +\infty$, then we will send $w \to +\infty$, and then $\varepsilon \to 0$. We shall reiterate this observation in the proof of Theorem 1.3.1. On the other hand, the $o_w(1)$ term is the same as the one in 2.4.2 and depends on the degree d of the polynomials, which will be fixed in applications.

Proof of Proposition 5.3.3. Fix an index $1 \leq i \leq k$ and consider a sufficiently large integer r. Using the mean value theorem and the fact that $|x_i'(t)|$ decreases to 0 faster than all fractional powers by Lemma 2.1.3, we deduce that

$$\max_{0\leq h\leq L(r)} |x_i(r + h) - x_i(r)| \leq L(r)|x_i'(r)| \lll 1.$$

In particular, there exists $\delta_0 > 0$ depending only on the functions $a_1, \dots, a_k$ and $L(t)$, such that

$$\max_{0\leq h\leq L(r)} |x_i(r + h) - x_i(r)| \ll r^{-\delta_0} \tag{5.26}$$

for all $1 \leq i \leq k$. Thus, we observe that if $\{p_i(n) + x_i(r)\} \in (\varepsilon, 1 - \varepsilon)$ and r is large enough in terms of $1/\varepsilon$, then we have that

$$\lfloor p_i(n) + x_i(n) \rfloor = \lfloor p_i(n) + x_i(r) \rfloor.$$

Naturally, we consider the set

$$\mathcal{B}_{i,r,\varepsilon} = \{n \in [r, r + L(r)]\colon \{p_i(n) + x_i(r)\} \in [0, \varepsilon] \cup [1 - \varepsilon, 1)\} \tag{5.27}$$

and take $\mathcal{B}_{r,\varepsilon} = \mathcal{B}_{1,r,\varepsilon} \cup \dots \cup \mathcal{B}_{k,r,\varepsilon}$. Now, we observe that the polynomial sequence p_i is well-distributed modulo 1, since it has at least one non-constant irrational coefficient. Therefore, if r is large enough, we have that the set $\mathcal{B}_{i,r,\varepsilon}$ has less than $3\varepsilon L(r)$ elements (say). Using the union bound, we conclude that the set $\mathcal{B}_{r,\varepsilon}$ has $O(\varepsilon k L(r))$ elements. This shows the first requirement of the proposition.

We have to establish (5.25). We shall set $M = \lfloor \varepsilon^{-1} \rfloor$ for brevity so that r is assumed to be very large in terms of M. Since the polynomials p_i have at least one non-constant irrational coefficient, we can use Weyl's criterion for well-distribution (see, for instance, [37, Theorem 5.2, Chapter 1]) to conclude that

$$\max_{1\leq m\leq M} \Big| \mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} e\big(m(p_i(n) + x_i(r))\big) \Big| = o_r(1),$$

for all r sufficiently large in terms of M, as we have assumed to be the case.[5] On the other hand, Lemma 2.4.2 implies that

$$\max_{1\leq m\leq M} \Big| \mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} \big(\Lambda_{w,b}(n) - 1\big) e\big(m(p_i(n) + x_i(r))\big) \Big| = o_w(1)$$

[4] See the notational conventions for the definition of $\lll$.

[5] A bound that is uniform over all $m \in \mathbb{N}$ is in general false, so we have to restrict m to a finite range.

for r sufficiently large in terms of w. Combining the last two bounds, we deduce that

$$\max_{1\le m\le M}\Big|\mathop{\mathbb{E}}_{r\le n\le r+L(r)}\Lambda_{w,b}(n)e\big(m(p_i(n)+x_i(r))\big)\Big| = o_w(1)+o_r(1). \tag{5.28}$$

Since we have estimates on the exponential sums weighted by $\Lambda_{w,b}(n)$, we can now make the passage to (5.25). To this end, we apply Theorem I for the probability measure

$$\nu(S)=\frac{\sum\limits_{r\le n\le r+L(r)}\Lambda_{w,b}(n)\delta_{\{p_i(n)+x_i(r)\}}(S)}{\sum\limits_{r\le n\le r+L(r)}\Lambda_{w,b}(n)}.^{6}$$

Setting

$$S_r=\sum_{r\le n\le r+L(r)}\Lambda_{w,b}(n)$$

for brevity, we conclude that

$$\frac{\sum\limits_{r\le n\le r+L(r)}\Lambda_{w,b}(n)\delta_{\{p_i(n)+x_i(r)\}}\big([0,\varepsilon]\cup[1-\varepsilon,1)\big)}{S_r}\ll 2\varepsilon+\frac{1}{M}+$$
$$\sum_{m=1}^{M}\frac{1}{m}\Big|\frac{1}{S_r}\sum_{r\le n\le r+L(r)}\Lambda_{w,b}(n)e\big(m(p_i(n)+x_i(r))\big)\Big|, \tag{5.29}$$

where the implied constant is absolute. Applying the bounds in (5.28) and recalling the definition of $\mathcal{B}_{i,r,\varepsilon}$, we conclude that

$$\sum_{r\le n\le r+L(r)}\Lambda_{w,b}(n)\mathbf{1}_{\mathcal{B}_{i,r,\varepsilon}}(n)\ll\Big(\varepsilon+\frac{1}{M}\Big)S_r+\sum_{m=1}^{M}\frac{L(r)}{m}(o_w(1)+o_r(1))$$
$$\ll\varepsilon S_r+L(r)\big(o_w(1)+o_r(1)\big)\log\frac{1}{\varepsilon}, \tag{5.30}$$

since $M=\lfloor\varepsilon^{-1}\rfloor$. Finally, we bound S_r by applying Corollary 2.4.3 to conclude that

$$S_r=\frac{\phi(W)}{W}\sum_{\substack{Wr+b\le n\le Wr+b+WL(r)\\ n\equiv b\ (W)}}\Lambda(n)\le\frac{\phi(W)}{W}\Big(\frac{2WL(r)\log r}{\phi(W)\log\big(\frac{L(r)}{W}\big)}+$$
$$O\Big(\frac{L(r)}{\log r}\Big)+O(r^{1/2}\log r)\Big)\ll L(r)(1+o_r(1)), \tag{5.31}$$

where we used the fact that $L(r)\gg t^{5/8}$ to bound the first fraction by an absolute constant. Applying this in (5.30), we conclude that

$$\frac{1}{L(r)}\sum_{r\le n\le r+L(r)}\Lambda_{w,b}(n)\mathbf{1}_{\mathcal{B}_{i,r,\varepsilon}}(n)\ll\varepsilon(1+o_r(1))+\big(o_w(1)+o_r(1)\big)\log\frac{1}{\varepsilon}.$$

Finally, we recall that $\mathcal{B}_{r,\varepsilon}=\mathcal{B}_{1,r,\varepsilon}\cup\dots\cup\mathcal{B}_{k,r,\varepsilon}$ and use the union bound to get

$$\frac{1}{L(r)}\sum_{r\le n\le r+L(r)}\Lambda_{w,b}(n)\mathbf{1}_{\mathcal{B}_{r,\varepsilon}}(n)\ll_k\varepsilon++o_w(1)\log\frac{1}{\varepsilon}+o_r(1),$$

provided that r is very large in terms of $1/\varepsilon, w$. This is the desired conclusion. □

[6] The denominator is non-zero if r is large enough.

5.3.2 Simultaneous approximation of Hardy field functions

In view of Proposition 5.3.1, we would like to show that we can find a function $L(t)$ such that the growth rate condition of the statement is satisfied for several functions in $\mathcal{H}$ simultaneously. This is the content of the following lemma. We will only need to consider the case where the functions dominate some fractional power, since for sub-fractional functions, we have Propositions 5.3.2 and 5.3.3 that can cover them adequately. We refer again to our notational conventions for the notation $\lll$.

Proposition 5.3.4. *Let $\ell \in \mathbb{N}$ and suppose $a_1, \ldots, a_\ell \in \mathcal{H}$ are strongly non-polynomial functions of polynomial growth that are not sub-fractional. Then, for all $0 < c < 1$, there exists a positive sub-linear function $L(t)$, such that $t^c \ll L(t) \ll t^{1-\varepsilon}$ for some $\varepsilon > 0$ and such that, for all $1 \leq i \leq \ell$, there exist positive integers k_i, which satisfy*

$$1 \lll \left|a_i^{(k_i)}(t)\right|^{-\frac{1}{k_i}} \lll L(t) \lll \left|a_i^{(k_i+1)}(t)\right|^{-\frac{1}{k_i+1}}.$$

Furthermore, the integers k_i can be chosen to be arbitrarily large, provided that c is sufficiently close to 1.

Proof. We will use induction on ℓ. For $\ell = 1$, it suffices to show that there exists a positive integer k, such that the function $\left|a^{(k+1)}(t)\right|^{-\frac{1}{k+1}}$ strongly dominates the function $\left|a^{(k)}(t)\right|^{-\frac{1}{k}}$. Then, we can pick the function $L(t)$ to be the geometric mean of these two functions to get our claim.[7]

Firstly, note that if we pick k sufficiently large, then we can ensure that $(a^{(k)}(t))^{-\frac{1}{k}} \gg t^c$, which would also imply the lower bound on the other condition imposed on the function $L(t)$. To see why this last claim is valid, observe that the derivatives of a satisfy the assumptions of Lemma 2.1.3, so that we have $a^{(k)}(t) \ll t^{-k}a(t)$. Thus, if d is a positive integer, such that t^d grows faster than $a(t)$ and we choose $k > \frac{d}{c} - 1$, we verify that our claim holds.

Secondly, we will show that for all $k \in \mathbb{N}$, we have

$$\left|a^{(k)}(t)\right|^{-\frac{1}{k}} \ll t^{1-\varepsilon}$$

for some $0 < \varepsilon < 1$, as this relation (with $k+1$ in place of k) will yield the upper bound on the growth of the function $L(t)$ that we chose above. This has been already established in point ii) of Lemma 3.5.1.

In order to complete the base case of the induction, we show that for all sufficiently large k, we have

$$\left|a^{(k)}(t)\right|^{-\frac{1}{k}} \lll \left|a^{(k)}(t)\right|^{-\frac{1}{k+1}}.$$

Equivalently, we prove that

$$\frac{\left|a^{(k+1)}(t)\right|^{-\frac{1}{k+1}}}{\left|a^{(k)}(t)\right|^{-\frac{1}{k}}} \gg t^{\delta} \tag{5.32}$$

for some $\delta > 0$ that will depend on k. This can be proven as the point i) in Lemma 3.5.1.

Assume that the claim has been established for the integer ℓ. Now, let $a_1, \ldots, a_{\ell+1}$ be functions that satisfy the hypotheses of the proposition. Our induction hypothesis implies that there exists a function $L(t)$ with $t^c \ll L(t) \ll t^{1-\varepsilon}$ and integers $k_1, \ldots, k_\ell$, such that

$$\left|a_i^{(k_i)}(t)\right|^{-\frac{1}{k_i}} \lll L(t) \lll \left|a_i^{(k_i+1)}(t)\right|^{-\frac{1}{k_i+1}}, \quad 1 \leq i \leq \ell.$$

Due to Proposition 2.1.5, there exists a positive integer s, such that

$$\left|a_{\ell+1}^{(s)}(t)\right|^{-\frac{1}{s}} \prec L(t) \prec \left|a_{\ell+1}^{(s+1)}(t)\right|^{-\frac{1}{s+1}}. \tag{5.33}$$

[7]It is straightforward to check that if $f \lll g$, then $f \lll \sqrt{fg} \lll g$, assuming, of course, that the square root is well-defined (e.g. when the functions f, g are eventually positive).

Without loss of generality, we may assume that c is sufficiently close to 1. This implies that the integer s can be chosen to be sufficiently large as well, so that the relation $|a_{\ell+1}^{(s)}(t)|^{-\frac{1}{s}} \lll |a_{\ell+1}^{(s+1)}(t)|^{-\frac{1}{s+1}}$ holds, as we established in the base case of the induction.

If each function strongly dominates the preceding one in (5.33), then we are finished. Therefore, assume that $L(t)$ is not strongly dominated by the function $|a_{\ell+1}^{(s+1)}(t)|^{-\frac{1}{s+1}}$ (the other case is similar). Note that for every $1 \leq i \leq \ell$, we have that

$$|a_i^{(k_i)}(t)|^{-\frac{1}{k_i}} \lll |a_{\ell+1}^{(s+1)}(t)|^{-\frac{1}{s+1}}.$$

Indeed, since the function $L(t)$ strongly dominates the function $|a_i^{(k_i)}(t)|^{-\frac{1}{k_i}}$ (by the induction hypothesis) and $L(t)$ grows slower than the the function $|a_{\ell+1}^{(s+1)}(t)|^{-\frac{1}{s+1}}$, this claim follows immediately. Among the functions $a_1, \ldots, a_{\ell+1}$, we choose a function for which the growth rate of $|a_i^{(k_i)}(t)|^{-\frac{1}{k_i}}$ is maximized.[8] Assume that this happens for the index $i_0 \in \{1, \ldots, \ell+1\}$ and observe that the function $|a_{\ell+1}^{(s+1)}(t)|^{-\frac{1}{s+1}}$ strongly dominates $|a_{i_0}^{(k_{i_0})}(t)|^{-\frac{1}{k_{i_0}}}$, because the first function grows faster than $L(t)$ and $L(t)$ strongly dominates the latter (in the case $i_0 = \ell+1$, this follows from the fact that $|a_{\ell+1}^{(s)}(t)|^{-\frac{1}{s}} \lll |a_{\ell+1}^{(s+1)}(t)|^{-\frac{1}{s+1}}$).

Define the function $\widetilde{L}(t)$ to be the geometric mean of the functions $|a_{i_0}^{(k_{i_0})}(t)|^{-\frac{1}{k_{i_0}}}$ and $|a_{\ell+1}^{(s+1)}(t)|^{-\frac{1}{s+1}}$. Observe that this function grows slower than the function $L(t)$, since it is strongly dominated by the function $|a_{\ell+1}^{(s+1)}(t)|^{-\frac{1}{s+1}}$, while the original function $L(t)$ is not. Due to its construction, we deduce that the function $\widetilde{L}(t)$ satisfies

$$|a_{\ell+1}^{(s)}(t)|^{-\frac{1}{s}} \lll \widetilde{L}(t) \lll |a_{\ell+1}^{(s+1)}(t)|^{-\frac{1}{s+1}}$$

and

$$|a_i^{(k_i)}(t)|^{-\frac{1}{k_i}} \lll \widetilde{L}(t)$$

for all $1 \leq i \leq \ell$. This is a simple consequence of the fact that $\widetilde{L}(t)$ strongly dominates the function $|a_{i_0}^{(k_{i_0})}(t)|^{-\frac{1}{k_{i_0}}}$ and the index i_0 was chosen so that the growth rate of the associated function is maximized. In addition, the function $L(t)$ grows faster than the function $\widetilde{L}(t)$, which implies that

$$\widetilde{L}(t) \prec L(t) \lll |a_i^{(k_i+1)}(t)|^{-\frac{1}{k_i+1}}$$

for all $1 \leq i \leq \ell$. The analogous relation in the case $i = \ell+1$ is also correct, as we pointed out previously. Therefore, the function $\widetilde{L}(t)$ satisfies all of our required properties and the induction is complete.

Finally, the assertion that the integers k_i can be made arbitrarily large follows by enlarging c appropriately and the fact that given a fixed $k_i \in \mathbb{N}$, the function $|a_i^{(k_i+1)}(t)|^{-\frac{1}{k_i+1}}$ cannot dominate all powers t^c with $c < 1$, as we displayed in the base case of the induction. □

We can actually weaken the hypothesis that the functions are strongly non-polynomial. The following proposition is more convenient to use and its proof is an immediate consequence of Proposition 5.3.4.

Proposition 5.3.5. *Let $\ell \in \mathbb{N}$ and suppose $a_1, \ldots, a_\ell \in \mathcal{H}$ are functions of polynomial growth, such that $|a_i(t) - p(t)| \ggg 1$, for all real polynomials $p(t)$ and every $i \in \{1, \ldots, \ell\}$. Then, for all $0 < c < 1$, there exists a positive sub-linear function $L(t)$, such that $t^c \prec L(t) \ll t^{1-\varepsilon}$ for some $\varepsilon > 0$ and such that there exist positive integers k_i, which satisfy*

$$1 \lll |a_i^{(k_i)}(t)|^{-\frac{1}{k_i}} \lll L(t) \lll |a_i^{(k_i+1)}(t)|^{-\frac{1}{k_i+1}}.$$

[8]In the case $i = \ell+1$, we are referring to the function $|a_i^{(s)}(t)|^{-\frac{1}{s}}$.

Proof. Each of the functions a_i can be written in the form $p_i(t)+x_i(t)$, where p_i is a polynomial with real coefficients and $x_i \in \mathcal{H}$ is strongly non-polynomial. The hypothesis implies that the functions x_i are not sub-fractional. If k is large enough, then we have $a_i^{(k)}(t)=x_i^{(k)}(t)$ for all $t\in\mathbb{R}$. The conclusion follows from Proposition 5.3.4 applied to the functions $x_i(t)$, where the corresponding integers k_i are chosen large enough so that the equality $a_i^{(k_i)}(t)=x_i^{(k_i)}(t)$ holds (that is, larger than the degrees of the polynomials $p_i(t)$). □

5.4 The main comparison

In this section, we will establish the main proposition that asserts that averages weighted by the W-tricked von-Mangoldt function are morally equal to the standard Cesàro averages over $\mathbb{N}$. In order to do this, we will use the polynomial approximations for our Hardy field functions and we will try to remove the error terms arising from these approximations using Propositions 5.3.1, 5.3.2 and 5.3.3. Firstly, we will use a lemma that allows us to pass from long averages over the interval $[1,N]$ to shorter averages over intervals of the form $[N, N+L(N)]$. This lemma is very similar to Lemma 3.2.2, the only difference being the presence of the unbounded weights.

Lemma 5.4.1. *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence in a normed space, such that $\|A_n\|\leq 1$ and let $L(t)\in\mathcal{H}$ be an (eventually) increasing sub-linear function, such that $L(t)\gg t^{\varepsilon}$ for some $\varepsilon>0$. Suppose that w is a fixed natural number. Then, we have*

$$\Big\|\mathop{\mathbb{E}}_{1\leq r\leq R}\big(\Lambda_{w,b}(r)-1\big)A_r\Big\| \leq \mathop{\mathbb{E}}_{1\leq r\leq R}\Big\|\mathop{\mathbb{E}}_{r\leq n\leq r+L(r)}\big(\Lambda_{w,b}(n)-1\big)A_n\Big\|+o_R(1),$$

uniformly for all $1\leq b\leq W$ with $(b,W)=1$.

Proof. Using the triangle inequality, we deduce that

$$\mathop{\mathbb{E}}_{1\leq r\leq R}\Big\|\mathop{\mathbb{E}}_{r\leq n\leq r+L(r)}\big(\Lambda_{w,b}(n)-1\big)A_n\Big\| \geq \Big\|\mathop{\mathbb{E}}_{1\leq r\leq R}\Big(\mathop{\mathbb{E}}_{r\leq n\leq r+L(r)}\big(\Lambda_{w,b}(n)-1\big)A_n\Big)\Big\|.$$

Therefore, our result will follow if we show that

$$\Big\|\mathop{\mathbb{E}}_{1\leq r\leq R}\Big(\mathop{\mathbb{E}}_{r\leq n\leq r+L(r)}\big(\Lambda_{w,b}(n)-1\big)A_n\Big)-\mathop{\mathbb{E}}_{1\leq r\leq R}\big(\Lambda_{w,b}(r)-1\big)A_r\Big\|=o_R(1).$$

Let u denote the inverse of the function $t+L(t)$, which is well-defined for sufficiently large t due to monotonicity. Furthermore, it is straightforward to derive that $\lim\limits_{t\to+\infty} u(t)/t=1$ from the fact that $t+L(t)$ also grows linearly. Now, we have

$$\mathop{\mathbb{E}}_{1\leq r\leq R}\Big(\mathop{\mathbb{E}}_{r\leq n\leq r+L(r)}\big(\Lambda_{w,b}(n)-1\big)A_n\Big)=\frac{1}{R}\Big(\sum_{n=1}^{R}p_R(n)\big(\Lambda_{w,b}(n)-1\big)A_n+ \sum_{n=R+1}^{R+L(R)}p_R(n)\big(\Lambda_{w,b}(n)-1\big)A_n\Big)$$

for some real numbers $p_R(n)$, which denote the number of appearances of A_n in the previous expression (weighted by the term $1/L(r)$ that appears on each inner average). Assuming that n (and thus R) is sufficiently large, so that $u(n)$ is positive, we can calculate $p_R(n)$ to be equal to

$$p_R(n)=\frac{1}{L(\lfloor u(n)\rfloor)+1}+\cdots+\frac{1}{L(n)+1}+o_n(1),$$

since the number A_n appears on the average $\mathop{\mathbb{E}}_{r\leq n\leq r+L(r)}$ if and only if $u(n)\leq r\leq n$. Note that $p_R(n)$ is actually independent of R (for n large enough) and therefore, we will denote it simply as $p(n)$ from now on. We have that

$$\lim_{n\to+\infty}p(n)=1. \tag{5.34}$$

This follows exactly as in the proof of Lemma 3.2.2.

Now, we show that

$$\frac{1}{R}\sum_{n=R+1}^{R+L(R)} p(n)\big(\Lambda_{w,b}(n)-1\big)A_n = o_R(1). \tag{5.35}$$

Bounding $p(n)$ trivially by 2 (since its limit is equal to 1) and $\|A_n\|$ by 1, we infer that it is sufficient to show that

$$\frac{1}{R}\sum_{n=R+1}^{R+L(R)} \big|\Lambda_{w,b}(n)-1\big| = o_R(1).$$

Using the triangle inequality and the fact that $L(r)\prec r$, this reduces to

$$\frac{1}{R}\sum_{n=R+1}^{R+L(R)} \Lambda_{w,b}(n) = o_R(1).$$

To establish this, we apply Corollary 2.4.3 to conclude that

$$\frac{1}{R}\sum_{n=R+1}^{R+L(R)} \frac{\phi(W)}{W}\Lambda(Wn+b) = \frac{1}{R}\sum_{\substack{WR+R+b\leq n\leq WR+R+b+WL(r)\\ n\equiv b\ (W)}} \Lambda(n) \leq$$
$$\frac{\phi(W)}{WR}\Big(\frac{2WL(R)\log R}{\phi(W)\log\big(\frac{L(R)}{W}\big)} + O\big(\frac{L(R)}{\log(WR+R+b)}\big) + O(R^{1/2}\log R)\Big) = o_R(1).$$

This follows from the fact that $L(R)\prec R$ and that the quantity $\log R/\log(L(R))$ is bounded by the hypothesis $L(R)\gg R^{\varepsilon}$.

In view of this, it suffices to show that

$$\Big\|\frac{1}{R}\sum_{n=1}^{R} p(n)\big(\Lambda_{w,b}(n)-1\big)A_n - \frac{1}{R}\sum_{n=1}^{R}\big(\Lambda_{w,b}(n)-1\big)A_n\Big\| = o_R(1).$$

We have

$$\Big\|\frac{1}{R}\sum_{n=1}^{R} p(n)\big(\Lambda_{w,b}(n)-1\big)A_n - \frac{1}{R}\sum_{n=1}^{R}\big(\Lambda_{w,b}(n)-1\big)A_n\Big\| \leq \frac{1}{R}\sum_{n=1}^{R}|p(n)-1||\Lambda_{w,b}(n)-1|,$$

by the triangle inequality. Now, given $\varepsilon>0$, we can bound this by

$$\frac{1}{R}\sum_{n=1}^{R}\varepsilon\big(\Lambda_{w,b}(n)+1\big) + o_R(1),$$

where the $o_R(1)$ term reflects the fact that the bound for $|p(n)-1|\leq\varepsilon$ is valid for large values of n only. It suffices to bound the term

$$\frac{\varepsilon}{R}\sum_{n=1}^{R}\Lambda_{w,b}(n),$$

since the remainder is simply $O(\varepsilon)$. However, using Corollary 2.4.3 (or the prime number theorem in arithmetic progressions), we see that this term is also $O(\varepsilon)$, exactly as we did above. Sending $\varepsilon\to 0$, we reach the desired conclusion. □

Proof of Theorem 1.3.1. We split this reduction into several steps. For a function $a\in\mathcal{H}$, we will use the notation $a_{w,b}(t)$ to denote the function $a(Wt+b)$ and we will need to keep in mind that the asymptotic constants must not depend on W and b. As is typical in these arguments, we shall rescale the functions $f_1,\ldots,f_\ell$ so that they are all bounded by 1.

Step 1: A preparatory decomposition of the functions

Each function a_{ij} can be written in the form

$$a_{ij}(t) = g_{ij}(t) + p_{ij}(t) + q_{ij}(t)$$

where $g_{ij}(t)$ is a strongly non-polynomial function (or identically zero), $p_{ij}(t)$ is either a polynomial with at least one non-constant irrational coefficient or a constant polynomial, and, lastly, $q_{ij}(t)$ is a polynomial with rational coefficients. Observe that there exists a fixed positive integer Q_0 for which all the polynomials $q_{ij}(Q_0 n + s_0)$ have integer coefficients except possibly the constant term, for all $0 \leq s_0 \leq Q_0$. These non-integer constant terms can be absorbed into the polynomial $p_{ij}(t)$. Therefore, splitting our average into the arithmetic progressions $(Q_0 n + s_0)$, it suffices to show that

$$\lim_{w\to+\infty} \limsup_{N\to+\infty} \max_{\substack{1\leq b\leq W \\ (b,W)=1}} \Big\| \frac{1}{N} \sum_{n=1}^{N} \big(\Lambda_{w,b}(Q_0 n + s_0) - 1\big) \prod_{j=1}^{\ell} \big(\prod_{i=1}^{k} T_i^{\lfloor a_{ij,w,b}(Q_0 n + s_0) \rfloor} \big) f_j \Big\|_{L^2(\mu)} = 0$$

for all $s_0 \in \{0, \dots, Q_0 - 1\}$. Observe that each one of the functions $a_{ij,w,b}(Q_0 t + s_0)$ satisfies either (1.19) or (1.20). Since the polynomials $q_{ij,w,b}(Q_0 n + s_0)$ have integer coefficients, we can rewrite the previous expression as

$$\lim_{w\to+\infty} \lim_{N\to+\infty} \max_{\substack{1\leq b\leq W \\ (b,W)=1}} \Big\| \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{s_0\ (Q_0)}(n) \big(\Lambda_{w,b}(n) - 1\big) \prod_{j=1}^{\ell} \big(\prod_{i=1}^{k} T_i^{\lfloor g_{ij,w,b}(n) + p_{ij,w,b}(n) \rfloor + q_{ij,w,b}(n)} \big) f_j \Big\|_{L^2(\mu)} = 0. \quad (5.36)$$

Step 2: Separating the iterates

Define the sets

$$S_1 = \{(i,j) \in [1,k] \times [1,\ell] \colon g_{ij}(t) \ll t^{\delta} \text{ for all } \delta > 0 \text{ and } p_{ij} \text{ is non-constant}\}, \quad (5.37)$$

and

$$S_2 = \{(i,j) \in [1,k] \times [1,\ell] \colon g_{ij}(t) \ll t^{\delta} \text{ for all } \delta > 0 \text{ and } p_{ij} \text{ is constant}\}, \quad (5.38)$$

whose union contains precisely the pairs (i,j), for which $g_{ij}(t)$ is sub-fractional.

Our first observation is that if a pair (i,j) belongs to S_2, then the function $a_{ij}(t)$ has the form $g_{ij}(t) + q_{ij}(t)$, where g_{ij} is sub-fractional and q_{ij} is a rational polynomial. Thus, (1.19) and (1.20) imply that we either have that $g_{ij}(t) \succ \log(t)$ or $g_{ij}(t)$ converges to a constant, as $t \to +\infty$. The constant can be absorbed into the constant polynomial p_{ij}. In view of this, we will subdivide S_2 further into the following two sets:

$$\begin{aligned} S_2' &= \{(i,j) \in S_2 \colon g_{ij}(t) \succ \log t\}, \\ S_2'' &= \{(i,j) \in S_2 \colon g_{ij}(t) \prec 1\}. \end{aligned} \quad (5.39)$$

Observe that iterates corresponding to pairs (i,j) that do not belong to the union $S_1 \cup S_2' \cup S_2''$ have an expression inside the integer part that has the form $g(t) + p(t)$, where g is a strongly non-polynomial function that is not sub-fractional. In particular, these functions satisfy the hypotheses of Proposition 5.3.5. Furthermore, functions that correspond to the set S_1 have the form $p(t) + x(t)$, where p is an irrational polynomial and x is sub-fractional, while functions in S_2' are sub-fractional functions that dominate $\log t$. We will use Proposition 5.3.3 and Proposition 5.3.2 for these two collections respectively. Finally, observe that if $(i,j) \in S_2''$, then for n sufficiently large, we can write

$$\lfloor a_{ij}(Q_0 n + s_0) \rfloor = q_{ij}(Q_0 n + s_0) + \lfloor c_{ij} \rfloor + e_{ij,Q_0 n + s_0},$$

where $e_{ij,Q_0 n+s_0} \in \{0, -1\}$ and c_{ij} is a constant term arising from the constant (in this case) polynomial p_{ij}. The error term $e_{ij,Q_0 n+s_0}$ actually exists only if c_{ij} is an integer. In particular, we have $e_{ij,Q_0 n+s_0} =$

0 for all large enough n when $g_{ij}(t)$ decreases to 0 and $e_{ij,Q_0n+s_0}=-1$ if $g_{ij}(t)$ increases to 0. Therefore, if we redefine the polynomials $q_{ij}(t)$ accordingly so that both $\lfloor c_{ij}\rfloor$ and the error term e_{ij,Q_0n+s_0} (which is independent of s_0) is absorbed into the constant term, we may assume without loss of generality that for all n sufficiently large, we have

$$\lfloor g_{ij}(Q_0n+s_0)+p_{ij}(Q_0n+s_0)\rfloor+q_{ij}(Q_0n+s_0)=q_{ij}(Q_0n+s_0).$$

We will employ this relation to simplify the iterates in (5.36), where n will be replaced by $Wn+b$.

We rewrite the limit in (5.36) as

$$\lim_{w\to+\infty}\ \limsup_{N\to+\infty}\ \max_{\substack{1\leq b\leq W\\(b,W)=1}}\Big\|\frac{1}{N}\sum_{n=1}^{N}\mathbf{1}_{s_0\ (Q_0)}(n)\big(\Lambda_{w,b}(n)-1\big)$$
$$\prod_{j=1}^{\ell}\Big(\prod_{i:\ (i,j)\in S_1}T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\cdot\prod_{i:\ (i,j)\in S_2'}T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\cdot$$
$$\prod_{i:\ (i,j)\in S_2''}T_i^{q_{ij,w,b}(n)}\cdot\prod_{i:\ (i,j)\notin S_1\cup S_2'\cup S_2''}T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\Big)f_j\Big\|_{L^2(\mu)}. \tag{5.40}$$

Step 3: Passing to short intervals

The functions $g_{ij}(t)+p_{ij}(t)$ with $(i,j)\in S_1$ satisfy the assumptions of Proposition 5.3.3, while the functions $g_{ij}(t)+p_{ij}(t)$ with $(i,j)\notin S_1\cup S_2'\cup S_2''$ satisfy the assumptions of Proposition 5.3.5 (thus, each one of them satisfies Proposition 5.3.1 for some appropriately chosen values of the integer k in that statement). Lastly, the functions of the set S_2' satisfy the assumptions of Propositions 5.3.2. It is straightforward to infer that, in each case, the corresponding property continues to hold when the functions $g_{ij}(t)+p_{ij}(t)$ are replaced by the functions $g_{ij,w,b}(t)+p_{ij,w,b}(t)$. This is a simple consequence of the fact that if $f\in\mathcal{H}$ has polynomial growth, then the functions f and $f_{w,b}$ have the same growth rate.

Let d_0 be the maximal degree appearing among the polynomials $p_{ij}(t)$. Then, we can find a sub-linear function $L(t)$ such that

$$t^{\frac{5}{8}}\lll L(t)\lll t \tag{5.41}$$

and, such that there exists positive integers k_{ij} for $(i,j)\notin S_1\cup S_2'\cup S_2''$, for which we have the growth inequalities

$$\left|g_{ij}^{(k_{ij})}(t)\right|^{-\frac{1}{k_{ij}}}\lll L(t)\lll\left|g_{ij}^{(k_{ij}+1)}(t)\right|^{-\frac{1}{k_{ij}+1}}. \tag{5.42}$$

Furthermore, we can assume that k_{ij} are very large compared to the maximal degree d_0 of the polynomials $p_{ij}(t)$, by taking $L(t)$ to grow sufficiently fast. We remark that (5.42) also implies the inequalities

$$\left|g_{ij,w,b}^{(k_{ij})}(t)\right|^{-\frac{1}{k_{ij}}}\lll L(t)\lll\left|g_{ij,w,b}^{(k_{ij}+1)}(t)\right|^{-\frac{1}{k_{ij}+1}}. \tag{5.43}$$

for any fixed w,b.

For the choice of $L(t)$ that we made above, we apply Lemma 5.4.1 to infer that it suffices to show that

$$\lim_{w\to+\infty}\ \limsup_{R\to+\infty}\ \max_{\substack{1\leq b\leq W\\(b,W)=1}}\ \mathop{\mathbb{E}}_{1\leq r\leq R}\Big\|\mathop{\mathbb{E}}_{r\leq n\leq r+L(r)}\mathbf{1}_{s_0\ (Q_0)}(n)\big(\Lambda_{w,b}(n)-1\big)$$
$$\prod_{j=1}^{\ell}\Big(\prod_{i:\ (i,j)\in S_1}T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\cdot\prod_{i:\ (i,j)\in S_2'}T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\cdot$$
$$\prod_{i:\ (i,j)\in S_2''}T_i^{q_{ij,w,b}(n)}\cdot\prod_{i:\ (i,j)\notin S_1\cup S_2'\cup S_2''}T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\Big)f_j\Big\|_{L^2(\mu)}=0. \tag{5.44}$$

Step 4: Reducing to polynomial iterates and using uniformity bounds

We now fix w (thus W) and the integer b. Suppose that R is sufficiently large and consider the expression

$$\mathcal{J}_{w,b,s_0}(R) := \mathop{\mathbb{E}}_{1\leq r\leq R} \Big\| \mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} \mathbf{1}_{s_0\ (Q_0)}(n)\big(\Lambda_{w,b}(n)-1\big) \prod_{j=1}^{\ell}\Big(\prod_{i:\ (i,j)\in S_1} T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor + q_{ij,w,b}(n)} \cdot \prod_{i:\ (i,j)\in S_2'} T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor + q_{ij,w,b}(n)} \cdot \prod_{i:\ (i,j)\in S_2''} T_i^{q_{ij,w,b}(n)} \cdot \prod_{i:\ (i,j)\notin S_1\cup S_2'\cup S_2''} T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor + q_{ij,w,b}(n)} \Big) f_j \Big\|_{L^2(\mu)}. \tag{5.45}$$

We will apply Propositions 5.3.1, 5.3.2 and 5.3.3 to replace the iterates with polynomials (with coefficients depending on r). Due to the nature of Proposition 5.3.2 (namely, that it excludes a small set of $r\in[1,R]$), we let $\mathcal{E}_{R,w,b}$ denote a subset of $\{1,\dots,R\}$, which will be constructed throughout the proof and will have small size. We remark that the iterates corresponding to S_2'' have been dealt with (morally), so we will focus our attention on the other three sets.

Let d be the maximum number among the degrees among the polynomials p_{ij}, q_{ij} and the integers k_{ij}. Let $\varepsilon>0$ be a small (but fixed) quantity and we assume that r is large enough in terms of $1/\varepsilon$, i.e., larger than some $R_0=R_0(\varepsilon)$. Observe that if R is sufficiently large, then we have $R_0\leq \varepsilon R$. We include the "small" r in the exceptional set $\mathcal{E}_{R,w,b}$, so that $\mathcal{E}_{R,w,b}$ now has at most εR elements. We will need to bound the expression $\mathcal{J}_{w,b,s_0}(R)$ for large R uniformly in b.

Throughout the rest of this step, we implicitly assume that all terms of the form $o_r(1)$ or $o_R(1)$ are allowed to depend on the parameters w and ε which will be fixed up until the end of Step 4. One can keep in mind the following hierarchy $\frac{1}{\varepsilon}\ll w\ll r$.

Case 1 : We first deal with the functions in S_2'. Fix an $(i,j)\in S_2'$ and consider the function $g_{ij,w,b}(n)+p_{ij,w,b}(n)$ appearing in the corresponding iterate. Observe that due to the definition of S_2' in (5.39), the polynomial $p_{ij}(t)$ is constant, so that $p_{ij,w,b}(t)$ is also constant. In addition, the function $g_{ij}(t)$ is a sub-fractional function and dominates $\log t$. Therefore, the same is true for the function $g_{ij,w,b}(t)$.

We apply Proposition 5.3.2: for all except at most εR values of $r\in[1,R]$, we have that

$$\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor = \lfloor g_{ij,w,b}(r)+p_{ij,w,b}(r)\rfloor \text{ for all } n\in[r,r+L(r)]. \tag{5.46}$$

For each $(i,j)\in S_2'$, we include the "bad" values of r to the set $\mathcal{E}_{R,w,b}$, so that the set $\mathcal{E}_{R,w,b}$ now has at most $(k\ell+1)\varepsilon R$ elements.

Case 2 : Now, we turn our attention to functions on the complement of the set $S_1\cup S_2'\cup S_2''$. The functions g_{ij} satisfy (5.43) and recall that we have chosen k_{ij} to be much larger than the degrees of the p_{ij}, so that the derivative of order k_{ij} of our polynomial vanishes. In conclusion, we may conclude that $g_{ij}(t)+p_{ij}(t)$ satisfies the assumptions of Proposition 5.3.1 for the integer k_{ij} (and the sub-linear function $L(t)$ that we have already chosen).

Given $A>0$, we infer that for all but $O_A(L(r)\log^{-A} r)$ values of $n\in[r,r+L(r)]$, we have

$$\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor = \lfloor \widetilde{p}_{ij,w,b,r}(n)\rfloor, \tag{5.47}$$

where $\widetilde{p}_{ij,w,b,r}(n)$ is the polynomial

$$\sum_{l=0}^{k_{ij}} \frac{(n-r)^l g^{(l)}_{ij,w,b}(r)}{l!} + p_{ij,w,b}(n).$$

Additionally, the polynomials $\widetilde{p}_{ij,w,b,r}$ satisfy

$$\frac{\big|\{n\in[r,r+L(r)]\colon \{\widetilde{p}_{ij,w,b,r}(n)\}\in[1-\delta,1)\}\big|}{L(r)} = \delta + O_A(\log^{-A} r) \tag{5.48}$$

for any $\delta < 1$. Practically, this last condition signifies that the polynomials $\widetilde{p}_{ij,w,b,r}$ satisfy the equidistribution condition in Proposition 5.2.3, which we shall invoke later.

Case 3 : Finally, we deal with the case of the set S_1. Proposition 5.3.3 suggests that there is a subset $\mathcal{B}_{w,b,r,\varepsilon}$ of $[r, r+L(r)]$ of size $O_{k,\ell}(\varepsilon L(r))$, such that for every $n \in [r, r+L(r)] \setminus \mathcal{B}_{w,b,r,\varepsilon}$, we have

$$\lfloor p_{ij,w,b}(n) + g_{ij,w,b}(n) \rfloor = \lfloor p_{ij,w,b}(n) + g_{ij,w,b}(r) \rfloor. \tag{5.49}$$

Additionally, the set $\mathcal{B}_{w,b,r,\varepsilon}$ satisfies

$$\frac{1}{L(r)} \sum_{r \le n \le r+L(r)} \Lambda_{w,b}(n) \mathbf{1}_{\mathcal{B}_{w,b,r,\varepsilon}}(n) \ll_{k,\ell,d} \varepsilon + o_w(1) \log \frac{1}{\varepsilon} + o_r(1). \tag{5.50}$$

We emphasize that the asymptotic constant in (5.50) depends only on k, l, d, so that the constant is the same regardless of the choice of the parameters w, b.

First of all, we apply (5.46) to simplify the expression for $\mathcal{J}_{w,b,s_0}(R)$. Namely, for any $r \notin \mathcal{E}_{R,w,b}$, we have that the inner average in the definition of $\mathcal{J}_{w,b,s_0}(R)$ is equal to

$$\Big\| \mathop{\mathbb{E}}_{r \le n \le r+L(r)} \mathbf{1}_{s_0\ (Q_0)}(n) \big(\Lambda_{w,b}(n) - 1\big) \prod_{j=1}^{\ell} \Big(\prod_{i:\ (i,j) \in S_1} T_i^{\lfloor g_{ij,w,b}(n) + p_{ij,w,b}(n) \rfloor + q_{ij,w,b}(n)} \cdot$$
$$\prod_{i:\ (i,j) \in S_2'} T_i^{\lfloor g_{ij,w,b}(r) + p_{ij,w,b}(r) \rfloor + q_{ij,w,b}(n)} \cdot \prod_{i:\ (i,j) \in S_2''} T_i^{q_{ij,w,b}(n)} \cdot$$
$$\prod_{i:\ (i,j) \notin S_1 \cup S_2' \cup S_2''} T_i^{\lfloor g_{ij,w,b}(n) + p_{ij,w,b}(n) \rfloor + q_{ij,w,b}(n)} \Big) f_j \Big\|_{L^2(\mu)}.$$

Thus, we have replaced the iterates of the set S_2' with polynomials in the averaging parameter n.

Secondly, we use (5.47) to deduce that for all, except at most $O_A(k\ell L(r) \log^{-A} r)$ values of $n \in [r, r+L(r)]$, the product of transformations appearing in the previous relation can be written as

$$\prod_{j=1}^{\ell} \Big(\prod_{i:\ (i,j) \in S_1} T_i^{\lfloor g_{ij,w,b}(n) + p_{ij,w,b}(n) \rfloor + q_{ij,w,b}(n)} \prod_{i:\ (i,j) \in S_2'} T_i^{\lfloor g_{ij,w,b}(r) + p_{ij,w,b}(r) \rfloor + q_{ij,w,b}(n)} \cdot$$
$$\prod_{i:\ (i,j) \in S_2''} T_i^{q_{ij,w,b}(n)} \cdot \prod_{i:\ (i,j) \notin S_1 \cup S_2' \cup S_2''} T_i^{\lfloor \widetilde{p}_{ij,w,b}(n) \rfloor + q_{ij,w,b}(n)} \Big) f_j. \tag{5.51}$$

The contribution of the exceptional set can be at most

$$k\ell \log(Wr + WL(r) + b) \cdot O_A(\log^{-A} r),$$

since each $\Lambda_{w,b}(n)$ is bounded by $\log(Wn + b)$. Therefore, if we choose $A \ge 2$, this contribution is $o_r(1)$ and we can rewrite the average over the corresponding short interval as

$$\Big\| \mathop{\mathbb{E}}_{r \le n \le r+L(r)} \mathbf{1}_{s_0\ (Q_0)}(n) \big(\Lambda_{w,b}(n) - 1\big) \prod_{j=1}^{\ell} \Big(\prod_{i:\ (i,j) \in S_1} T_i^{\lfloor g_{ij,w,b}(n) + p_{ij,w,b}(n) \rfloor + q_{ij,w,b}(n)}$$
$$\prod_{i:\ (i,j) \in S_2'} T_i^{\lfloor g_{ij,w,b}(r) + p_{ij,w,b}(r) \rfloor + q_{ij,w,b}(n)} \cdot \prod_{i:\ (i,j) \in S_2''} T_i^{q_{ij,w,b}(n)} \cdot$$
$$\prod_{i:\ (i,j) \notin S_1 \cup S_2' \cup S_2''} T_i^{\lfloor \widetilde{p}_{ij,w,b,r}(n) \rfloor + q_{ij,w,b}(n)} \Big) f_j \Big\|_{L^2(\mu)} + o_r(1). \tag{5.52}$$

Thus, we have reduced our iterates to polynomial form in this case as well.

Finally, we follow the same procedure for the set S_1. Namely, for all integers n in the interval $[r, r+L(r)]$ such that $n \notin \mathcal{B}_{w,b,r,\varepsilon}$, we use (5.49) to rewrite (5.51) as

$$\prod_{j=1}^{\ell}\Big(\prod_{i\colon (i,j)\in S_1} T_i^{\lfloor g_{ij,w,b}(r)+p_{ij,w,b}(n)\rfloor + q_{ij,w,b}(n)} \prod_{i\colon (i,j)\in S_2'} T_i^{\lfloor g_{ij,w,b}(r)+p_{ij,w,b}(r)\rfloor + q_{ij,w,b}(n)}\cdot$$
$$\prod_{i\colon (i,j)\in S_2''} T_i^{q_{ij,w,b}(n)}\cdot \prod_{i\colon (i,j)\notin S_1\cup S_2'\cup S_2''} T_i^{\lfloor \widetilde{p}_{ij,w,b}(n)\rfloor + q_{ij,w,b}(n)}\Big) f_j.$$

The contribution of the set $\mathcal{B}_{w,b,r,\varepsilon}$ on the average over the interval $[r, r+L(r)]$ can be estimated using the triangle inequality. More specifically, this contribution is smaller than

$$\frac{1}{L(r)}\sum_{r\leq n\leq r+L(r)} \mathbf{1}_{s_0\ (Q_0)}(n)\Big|\Lambda_{w,b}(n)-1\Big|\mathbf{1}_{\mathcal{B}_{w,b,r,\varepsilon}}(n).$$

We bound the characteristic function $\mathbf{1}_{s_0\ (Q_0)}$ trivially by 1, so that the above quantity is smaller than

$$\frac{1}{L(r)}\sum_{r\leq n\leq r+L(r)} \Lambda_{w,b}(n)\mathbf{1}_{\mathcal{B}_{w,b,r,\varepsilon}}(n) + \frac{1}{L(r)}\sum_{r\leq n\leq r+L(r)} \mathbf{1}_{\mathcal{B}_{w,b,r,\varepsilon}}(n). \tag{5.53}$$

The second term contributes $O_{k,\ell}(\varepsilon)$, since $\mathcal{B}_{w,b,r,\varepsilon}$ has at most $O_{k,\ell}(\varepsilon L(r))$ elements. On the other hand, we have a bound for the first term already in (5.50). Thus, the total contribution is $O_{k,\ell,d}(1)$ times the expression

$$\varepsilon + o_w(1)\log\frac{1}{\varepsilon} + o_r(1).$$

In view of the above, we deduce that the average in (5.52) is bounded by $O_{k,\ell,d}(1)$ times

$$\Big\|\mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} \mathbf{1}_{s_0(Q_0)}(n)\big(\Lambda_{w,b}(n)-1\big)\prod_{j=1}^{\ell}\Big(\prod_{i\colon (i,j)\in S_1} T_i^{\lfloor g_{ij,w,b}(r)+p_{ij,w,b}(n)+q_{ij,w,b}(n)\rfloor}$$
$$\prod_{i\colon (i,j)\in S_2'} T_i^{\lfloor g_{ij,w,b}(r)+p_{ij,w,b}(r)+q_{ij,w,b}(n)\rfloor}\cdot \prod_{i\colon (i,j)\in S_2''} T_i^{\lfloor q_{ij,w,b}(n)\rfloor}\cdot$$
$$\prod_{i\colon (i,j)\notin S_1\cup S_2'\cup S_2''} T_i^{\lfloor \widetilde{p}_{ij,w,b}(n)+q_{ij,w,b}(n)\rfloor}\Big) f_j\Big\|_{L^2(\mu)} + \varepsilon + o_w(1)\log\frac{1}{\varepsilon} + o_r(1). \tag{5.54}$$

Here, we moved the polynomials $q_{ij,w,b}$ back inside the integer parts, which we are allowed to do since they have integer coefficients.

The polynomials in the iterates corresponding to S_1, S_2', S_2'', and the complement of $S_1\cup S_2'\cup S_2''$ fulfill the hypothesis of Proposition 5.2.3. To keep the number of parameters lower, we will apply this proposition for $\delta=\varepsilon$, where we have assumed that ε is a very small parameter. Accordingly, we assume (as we may) that w and r are much larger than $\frac{1}{\varepsilon}$. To see why the hypotheses are satisfied, observe that for the first set, this follows from the fact that $p_{ij,w,b}$ has at least one non-constant irrational coefficient (since p_{ij} is non-constant by the definition of S_1). Therefore, the number of integers $n\in[r, r+L(r)]$ for which we have

$$\{g_{ij,w,b}(r)+p_{ij,w,b}(n)+q_{ij,w,b}(n)\}\in(1-\varepsilon, 1)$$

is smaller than $2\varepsilon L(r)$ for r sufficiently large. At the same time, the result is immediate for the second and third sets, since the iterates involve polynomials with integer coefficients (except, possibly, their constant terms). For the final set, this claim follows from (5.48).

In view of the prior discussion, we conclude that there exists a positive integer s, that depends only on d, k, ℓ, such that the expression in (5.54) is bounded by

$$\varepsilon^{-k\ell}\big\|\mathbf{1}_{s_0\ (Q_0)}\big(\Lambda_{w,b}(n)-1\big)\big\|_{U^s(r,r+sL(r)]} + \varepsilon^{-k\ell}o_w(1) + o_\varepsilon(1)(1+o_w(1)) +$$
$$\varepsilon + o_w(1)\log\frac{1}{\varepsilon} + o_r(1). \tag{5.55}$$

Applying Lemma 2.2.2, we can bound the previous Gowers norm along the residue class s_0 (Q_0) as follows:

$$\big\|\mathbf{1}_{s_0\ (Q_0)}\big(\Lambda_{w,b}(n)-1\big)\big\|_{U^s(r,r+sL(r)]} \leq \big\|\Lambda_{w,b}(n)-1\big\|_{U^s(r,r+sL(r)]}. \tag{5.56}$$

In view of the arguments above, we conclude that, for every $r\notin \mathcal{E}_{R,w,b}$, the following inequality holds

$$\Big\| \mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} \mathbf{1}_{s_0(Q_0)}(n)\big(\Lambda_{w,b}(n)-1\big) \prod_{j=1}^{\ell}\Big(\prod_{i:\,(i,j)\in S_1} T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\cdot \prod_{i:\,(i,j)\in S_2'} T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\cdot \prod_{i:\,(i,j)\in S_2''} T_i^{q_{ij,w,b}(n)}\cdot \prod_{i:\,(i,j)\notin S_1\cup S_2'\cup S_2''} T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\Big) f_j\Big\|_{L^2(\mu)} \ll_{k,\ell,d}$$
$$\varepsilon^{-k\ell}\big\|\big(\Lambda_{w,b}(n)-1\big)\big\|_{U^s(r,r+sL(r)]}+\varepsilon+\big(\varepsilon^{-k\ell}+\log\frac{1}{\varepsilon}+o_\varepsilon(1)\big)o_w(1)+o_\varepsilon(1)+o_r(1).$$

We apply this estimate to the double average defining $\mathcal{J}_{w,b,s_0}(R)$ in (5.45). This estimate holds for every $r\notin\mathcal{E}_{R,w,b}$ and, thus, we need an estimate for the values of r in this exceptional set. In order to achieve this, we recall that the set $\mathcal{E}_{R,w,b}$ has at most $(2k\ell+1)\varepsilon R$ elements. For each $r\in\mathcal{E}_{R,w,b}$, we use the triangle inequality to bound the average over the corresponding short interval by

$$\frac{1}{L(r)}\sum_{\substack{r\leq n\leq r+L(r)\\ n\equiv s_0\ (Q_0)}}(\Lambda(Wn+b)+1).$$

We bound the characteristic function of the residue class $n\equiv s_0$ (Q_0) trivially by 1 and apply Corollary 2.4.3 to conclude that this expression is $O(1)+o_r(1)$, using similar estimates as the ones used in the proof of Proposition 5.3.3 (see (5.31)). Therefore, the contribution of the set $\mathcal{E}_{R,w,b}$ is at most $O_{k,\ell}(\varepsilon)+o_r(1)$. Combining all of the above, we arrive at the estimate

$$\mathcal{J}_{w,b,s_0}(R)\ll_{d,k,\ell}\ \varepsilon^{-k\ell}\Big(\mathop{\mathbb{E}}_{1\leq r\leq R}\big\|\big(\Lambda_{w,b}(n)-1\big)\big\|_{U^s(r,r+sL(r)]}\Big)+\varepsilon^{-k\ell}o_w(1)+ o_\varepsilon(1)(1+o_w(1))+o_R(1). \tag{5.57}$$

We restate (5.44) here. Namely, we want to show that

$$\limsup_{R\to+\infty}\max_{\substack{1\leq b\leq W\\ (b,W)=1}}\mathcal{J}_{w,b,s_0}(R)=o_w(1).$$

Applying (5.57), we conclude that for a fixed w, we have

$$\limsup_{R\to+\infty}\max_{\substack{1\leq b\leq W\\ (b,W)=1}}\mathcal{J}_{w,b,s_0}(R)\ll_{d,k,\ell}\varepsilon^{-k\ell}\Big(\lim_{R\to+\infty}\mathop{\mathbb{E}}_{1\leq r\leq R}\max_{\substack{1\leq b\leq W\\ (b,W)=1}}\big\|\big(\Lambda_{w,b}(n)-1\big)\big\|_{U^s(r,r+L(r)]}\Big)+ \varepsilon^{-k\ell}o_w(1)+o_\varepsilon(1)(1+o_w(1)).$$

Due to Theorem K, we have that

$$\max_{\substack{1\leq b\leq W\\ (b,W)=1}}\big\|\big(\Lambda_{w,b}(n)-1\big)\big\|_{U^s(r,r+L(r)]}=o_w(1)$$

for every sufficiently large r. Thus, we conclude that

$$\limsup_{R\to+\infty}\max_{\substack{1\leq b\leq W\\ (b,W)=1}}\mathcal{J}_{w,b,s_0}(R)\ll_{d,k,\ell}\varepsilon^{-k\ell}o_w(1)+o_\varepsilon(1)(1+o_w(1)).$$

Step 5: Putting all the bounds together

We restate here our conclusion. We have shown that for all fixed integers w and real number $0<\varepsilon<1$, we have

$$\limsup_{R\to+\infty}\lim_{N\to+\infty}\max_{\substack{1\leq b\leq W\\(b,W)=1}}\Big\|\frac{1}{N}\sum_{n=1}^{N}\mathbf{1}_{s_0(Q_0)}(n)\big(\Lambda_{w,b}(n)-1\big)\prod_{j=1}^{\ell}\big(\prod_{i=1}^{k}T_i^{\lfloor a_{ij,w,b}(n)\rfloor}\big)f_j\Big\|_{L^2(\mu)}$$
$$\ll_{d,k,\ell}\varepsilon^{-k\ell}o_w(1)+o_\varepsilon(1)(1+o_w(1)),\quad (5.58)$$

where we recall that d was the maximum among the integers k_{ij} and the degrees of the polynomials p_{ij},q_{ij} (all of these depend only on the initial functions a_{ij}). Sending $w\to+\infty$, we deduce that the limit in (5.36) (in view of (5.58)) is smaller than a constant (depending on k,ℓ,d) multiple of $o_\varepsilon(1)$. Sending $\varepsilon\to 0$, we conclude that the original limit is 0, which is the desired result. □

5.5 Proofs of the remaining theorems

We finish the proofs of our theorems in this section.

5.5.1 Proof of the convergence results

Proof of Theorem 1.3.2. Let $(X,\mathcal{X},\mu,T_1,\ldots,T_k)$ be the system and $a_{ij}\in\mathcal{H}$ the functions in the statement. In view of Lemma 2.4.1, it suffices to show that the averages

$$A(N):=\frac{1}{N}\sum_{n=1}^{N}\Lambda(n)\big(\prod_{i=1}^{k}T_i^{\lfloor a_{i1}(n)\rfloor}\big)f_1\cdot\ldots\cdot\big(\prod_{i=1}^{k}T_i^{\lfloor a_{i\ell}(n)\rfloor}\big)f_\ell$$

converge in $L^2(\mu)$. For a fixed $w\in\mathbb{N}$, we define $W=\prod_{p\leq w,p\in\mathbb{P}}p$ as usual and let $b\in\mathbb{N}$. We define

$$B_{w,b}(N):=\frac{1}{N}\sum_{n=1}^{N}\big(\prod_{i=1}^{k}T_i^{\lfloor a_{i1}(Wn+b)\rfloor}\big)f_1\cdot\ldots\cdot\big(\prod_{i=1}^{k}T_i^{\lfloor a_{i\ell}(Wn+b)\rfloor}\big)f_\ell.$$

Let $\varepsilon>0$. Using Theorem 1.3.1, we can find $w_0\in\mathbb{N}$ (which yields a corresponding W_0) such that

$$\Big\|A(W_0N)-\frac{1}{\phi(W_0)}\sum_{\substack{1\leq b\leq W_0\\(b,W_0)=1}}B_{w_0,b}(N)\Big\|_{L^2(\mu)}=O(\varepsilon)\quad (5.59)$$

for all N sufficiently large. Our hypothesis implies that the sequence of bounded functions $B_{w_0,b}(N)$ is a Cauchy sequence in $L^2(\mu)$, which, in conjunction with (5.59), implies that the sequence $A(W_0N)$ is a Cauchy sequence. In particular, we have

$$\|A(W_0M)-A(W_0N)\|_{L^2(\mu)}=O(\varepsilon),$$

for all N,M sufficiently large. Finally, since

$$\|A(W_0N+b)-A(W_0N)\|_{L^2(\mu)}=o_N(1),$$

for all $1\leq b\leq W_0$, we conclude that $A(N)$ is a Cauchy sequence, which implies the required convergence.

Furthermore, if the sequence $B_{w,b}(N)$ converges to the function F in $L^2(\mu)$ for all $w,r\in\mathbb{N}$, then (5.59) implies that $\|A(W_0N)-F\|_{L^2(\mu)}=O(\varepsilon)$, for all large enough N. Repeating the same argument as above, we infer that $A(N)$ converges to the function F in norm, as we desired. □

Proof of Theorem 1.3.3. Let $a \in \mathcal{H}$ satisfy either (1.23) or (1.24), $k \in \mathbb{N}$, $(X, \mathcal{X}, \mu, T)$ be any measure-preserving system, and functions $f_1, \ldots, f_k \in L^\infty(\mu)$. Observe that in either case, the function a satisfies (1.19) or (1.20). In addition, when $a(t)$ satisfies either of the two latter conditions, then the function $a(Wt + b)$ satisfies the same condition, for all $W, b \in \mathbb{N}$.

Using [13, Theorem 2.1],[9] we have that, for all $W, b \in \mathbb{N}$, the averages

$$\frac{1}{N}\sum_{n=1}^{N} T^{\lfloor a(Wn+b)\rfloor} f_1 \cdot \ldots \cdot T^{k\lfloor a(Wn+b)\rfloor} f_k$$

converge in $L^2(\mu)$. We conclude that the two conditions of Theorem 1.3.2 are satisfied, which shows that the desired averages converge.

In particular, if a satisfies condition (1.23), we can invoke [13, Theorem 2.2] to conclude that the limit of the averages

$$\frac{1}{N}\sum_{n=1}^{N} T^{\lfloor a(Wn+b)\rfloor} f_1 \cdot \ldots \cdot T^{k\lfloor a(Wn+b)\rfloor} f_k$$

is equal to the limit (in $L^2(\mu)$) of the averages

$$\frac{1}{N}\sum_{n=1}^{N} T^{n} f_1 \cdot \ldots \cdot T^{kn} f_k.$$

Again, Theorem 1.3.2 yields the desired conclusion. □

Proof of Theorem 1.3.4. We work analogously as in the proof of Theorem 1.3.3. The only difference is that in this case, we use Theorem 1.1.1 to deduce that, for all $W \in \mathbb{N}$, $b \in \mathbb{N}$ positive integers W and b, the averages

$$\frac{1}{N}\sum_{n=1}^{N} T^{\lfloor a_1(Wn+b)\rfloor} f_1 \cdot \ldots \cdot T^{\lfloor a_k(Wn+b)\rfloor} f_k$$

converge in $L^2(\mu)$ to the product $\widetilde{f}_1 \cdot \ldots \cdot \widetilde{f}_k$. The result follows from Theorem 1.3.2. □

Proof of Theorem 1.3.5. The proof follows identically as the one of Theorem 1.3.4 by using [14, Theorem 2.3] instead of Theorem 1.1.1. □

5.5.2 Proof of the recurrence results

We prove here the multiple recurrence theorems. In view of the Furstenberg correspondence principle, the corollaries follow easily.

Proof of Theorem 1.3.6. (a) We apply Theorem 1.3.3 for the functions $f_1 = \cdots = f_k = \mathbf{1}_A$. Since convergence in $L^2(\mu)$ implies weak convergence, integrating along A the relation

$$\lim_{N\to+\infty} \frac{1}{\pi(N)} \sum_{p\in\mathbb{P}:\, p\leq N} T^{\lfloor a(p)\rfloor}\mathbf{1}_A \cdot \ldots \cdot T^{k\lfloor a(p)\rfloor}\mathbf{1}_A = \lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} T^{n}\mathbf{1}_A \cdot \ldots \cdot T^{kn}\mathbf{1}_A,$$

and applying Furstenberg's multiple recurrence theorem we infer that

$$\lim_{N\to+\infty} \frac{1}{\pi(N)} \sum_{p\in\mathbb{P}:\, p\leq N} \mu\big(A \cap T^{-\lfloor a(p)\rfloor}A \cap \cdots \cap T^{-k\lfloor a(p)\rfloor}A\big) > 0,$$

which is the desired result.

[9] There is a slight technicality here, in that we would need the assumption that the function $a(Wn+b)$ belongs to $\mathcal{H}$ in order to apply Theorem 2.2 from [13], However, the proof in [13] only requires some specific growth conditions on the derivatives of the function $a(Wn+b)$ (specifically those outlined in equation 26 of that paper), which follow naturally from the assumption that $a \in \mathcal{H}$.

(b) We write $a(t) = cq(t) + \varepsilon(t)$, where $q(t) \in \mathbb{Z}[t]$, $q(0) = 0$, $c \in \mathbb{R}$ and $\varepsilon(t)$ is a function that converges to 0, as $t \to +\infty$. Using [33, Proposition 3.8], we have that there exists c_0 depending only on $\mu(A)$, the degree of q and k, such that

$$\liminf_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} \mu(A \cap T^{-[[cq(n)]]}A \cap \cdots \cap T^{-k[[cq(n)]]}A) \geq c_0.$$

Now, we consider two separate cases. If c is rational with denominator Q in lowest terms, then for t sufficiently large, we have $|\varepsilon(t)| \leq (2Q)^{-1}$. Therefore, we immediately deduce that

$$[[cq(t) + \varepsilon(t)]] = [[cq(t)]].$$

Thus, we conclude that

$$\liminf_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} \mu(A \cap T^{-[[cq(n)+\varepsilon(n)]]}A \cap \cdots \cap T^{-k[[cq(n)+\varepsilon(n)]]}A) \geq c_0. \tag{5.60}$$

If c is irrational, then the polynomial $cq(t)$ is uniformly distributed mod 1. Given $\delta > 0$, we consider the set $S := \{n \in \mathbb{N}\colon \{cq(n)\} \in [\delta, 1-\delta]\}$, which has density $1 - 2\delta$. Therefore, we have

$$\Big|\frac{1}{N}\sum_{n=1}^{N} \mu(A \cap T^{-[[cq(n)+\varepsilon(n)]]}A \cap \cdots \cap T^{-k[[cq(n)+\varepsilon(n)]]}A) - \frac{1}{N}\sum_{n=1}^{N} \mu(A \cap T^{-[[cq(n)]]}A \cap \cdots \cap T^{-k[[cq(n)]]}A)\Big| \leq 2\delta + o_N(1).$$

Sending $\delta \to 0^+$, we derive (5.60) in this case as well.

Notice that since c_0 depends only on the degree of q, we have that

$$\liminf_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} \mu(A \cap T^{-[[cq(Rn)+\varepsilon(Rn)]]}A \cap \cdots \cap T^{-k[[cq(Rn)+\varepsilon(Rn)]]}A) \geq c_0,$$

for all positive integers R. Now, we apply Theorem 1.3.1 with $b = 1$ and the functions $a(\cdot - 1)$, where we recall that $a(t) = cq(t) + \varepsilon(t)$ to obtain that for some sufficiently large w, we have

$$\liminf_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} \Lambda_{w,1}(n)\mu\big(A \cap T^{-\lfloor a(Wn)\rfloor}A \cap \cdots \cap T^{-k\lfloor a(Wn)\rfloor}A\big) \geq c_0/2,$$

where W is defined as usual in terms of w. Finally, we observe that we can replace the function $\Lambda(n)$ in the previous relation with the function $\Lambda(n)\mathbf{1}_{\mathbb{P}}(n)$ since the contribution of the prime powers (i.e. with exponent ≥ 2) is negligible on the average. Therefore, we conclude that

$$\liminf_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} \Lambda_{w,1}(n)\mathbf{1}_{\mathbb{P}}(Wn+1)\mu\big(A \cap T^{-\lfloor a(Wn)\rfloor}A \cap \cdots \cap T^{-k\lfloor a(Wn)\rfloor}A\big) \geq c_0/2,$$

which implies the desired result. Analogously, we reach the expected conclusion for the set $\mathbb{P} + 1$ instead of $\mathbb{P} - 1$. ☐

Proof of Theorem 1.3.8. Similarly to the proof of Theorem 1.3.6, we apply Theorem 1.3.4 for the functions $f_1 = \cdots = f_k = \mathbf{1}_A$. We deduce that

$$\lim_{N\to+\infty} \frac{1}{\pi(N)} \sum_{p\in\mathbb{P}\colon p\leq N} \mu\big(A \cap T^{-\lfloor a_1(p)\rfloor}A \cap \cdots \cap T^{-\lfloor a_k(p)\rfloor}A\big) = \int \mathbf{1}_A \cdot \big(\mathbb{E}(\mathbf{1}_A|\mathcal{I}(T))\big)^k \, d\mu. \tag{5.61}$$

However, using that the function $\mathbf{1}_A$ is non-negative and Hölder's inequality, we get

$$\int \mathbf{1}_A \cdot \big(\mathbb{E}(\mathbf{1}_A|\mathcal{I}(T))\big)^k \, d\mu \geq \Big(\int \mathbb{E}(\mathbf{1}_A|\mathcal{I}(T))\, d\mu\Big)^{k+1} = \big(\mu(A)\big)^{k+1},$$

and the conclusion follows. ☐

Proof of Theorem 1.3.10. The proof is similar to the proof of Theorem 1.3.8. The only distinction is made in (5.61), namely we have

$$\lim_{N\to+\infty}\frac{1}{\pi(N)}\sum_{p\in\mathbb{P}\colon p\le N}\mu\big(A_0\cap T_1^{-\lfloor a_1(p)\rfloor}A_1\cap\cdots\cap T_k^{-\lfloor a_k(p)\rfloor}A_k\big)=$$

$$\int \mathbf{1}_{A_0}\cdot\mathbb{E}(\mathbf{1}_{A_1}|\mathcal{I}(T_1))\cdot\ldots\cdot\mathbb{E}(\mathbf{1}_{A_k}|\mathcal{I}(T_k))\,d\mu,$$

where the sets $A_0, A_1, \ldots, A_k$ satisfy the hypothesis. Since each function $\mathbb{E}(\mathbf{1}_{A_i}|\mathcal{I}(T_i))$ is T_i-invariant, we deduce that the integral on the right-hand side is larger than

$$\int f\cdot\mathbb{E}(f|\mathcal{I}(T_1))\cdot\ldots\cdot\mathbb{E}(f|\mathcal{I}(T_k))\,d\mu,$$

where $f=\mathbf{1}_{A_0\cap T^{\ell_1}A_1\cap\cdots\cap T^{\ell_k}A_k}$. However, since the function f is non-negative, [8, Lemma 1.6] implies that

$$\int f\cdot\mathbb{E}(f|\mathcal{I}(T_1))\cdot\ldots\cdot\mathbb{E}(f|\mathcal{I}(T_k))\,d\mu\ge\Big(\int f\,d\mu\Big)^{k+1}=\mu(A)^{k+1},$$

and the conclusion follows. □

5.5.3 Proof of the equidistribution results in nilmanifolds

In this final part of this section, we offer a proof for Theorem 1.3.12. The main tool is the approximation of Lemma 2.3.9.

Proof of Theorem 1.3.12. Let X and $g_1, \ldots, g_k, x_1, \ldots, x_k$ be as in the statementthe section?, we offer a proof for Theorem 1.12. The main tool is the approximation of and let s denote the nilpotency degree of X. It suffices to show that, for any continuous functions $f_1, \ldots, f_s$ on X, we have the following:

$$\lim_{N\to+\infty}\frac{1}{\pi(N)}\sum_{p\in\mathbb{P}\colon p\le N}f_1(g_1^{\lfloor a_1(p)\rfloor}x_1)\cdot\ldots\cdot f_k(g_k^{\lfloor a_k(p)\rfloor}x_k)=\int_{Y_1}f_1\,dm_{Y_1}\cdot\ldots\cdot\int_{Y_k}f_k\,dm_{Y_k},$$

where $Y_i=\overline{(g_i^{\mathbb{Z}}x_i)}$ for all admissible values of i. We rewrite this in terms of the von Mangoldt function as

$$\lim_{N\to+\infty}\frac{1}{N}\sum_{n=1}^{N}\Lambda(n)f_1(g_1^{\lfloor a_1(n)\rfloor}x_1)\cdot\ldots\cdot f_k(g_k^{\lfloor a_k(n)\rfloor}x_k)=\int_{Y_1}f_1\,dm_{Y_1}\cdot\ldots\cdot\int_{Y_k}f_k\,dm_{Y_k},\tag{5.62}$$

where the equivalence of the last two relations is a consequence of Lemma 2.4.1.

Our equidistribution assumption implies that for all $W, b\in\mathbb{N}$, we have

$$\lim_{N\to+\infty}\frac{1}{N}\sum_{n=1}^{N}f_1(g_1^{\lfloor a_1(Wn+b)\rfloor}x_1)\cdot\ldots\cdot f_k(g_k^{\lfloor a_k(Wn+b)\rfloor}x_k)=\int_{Y_1}f_1\,dm_{Y_1}\cdot\ldots\cdot\int_{Y_k}f_k\,dm_{Y_k}.\tag{5.63}$$

We write $Y_i=G_i/\Gamma_i$ for some nilpotent Lie groups G_i with discrete and co-compact subgroups Γ_i and denote $Y=Y_1\times\cdots\times Y_k$. Define the function $F\colon Y\to\mathbb{C}$ by $F(y_1,\ldots,y_k)=f_1(y_1)\cdot\ldots\cdot f_k(y_k)$ and rewrite (5.62) as

$$\lim_{N\to+\infty}\frac{1}{N}\sum_{n=1}^{N}\Lambda(n)F(\widetilde{g}_1^{\lfloor a_1(n)\rfloor}\cdot\ldots\cdot\widetilde{g}_k^{\lfloor a_k(n)\rfloor}\widetilde{x})=\int_{Y}F\,dm_Y,\tag{5.64}$$

where $\widetilde{g}_i$ is the element on the nilpotent Lie group $G_1\times\cdots\times G_k$ whose i-th coordinate is equal to g_i and the rest of its entries are the corresponding identity elements. Lastly, $\widetilde{x}$ is the point $(x_1,\ldots,x_k)\in Y$. Similarly, we rewrite (5.63) as

$$\lim_{N\to+\infty}\frac{1}{N}\sum_{n=1}^{N}F(\widetilde{g}_1^{\lfloor a_1(Wn+b)\rfloor}\cdot\ldots\cdot\widetilde{g}_k^{\lfloor a_k(Wn+b\rfloor}\widetilde{x})=\int_{Y}F\,dm_Y.\tag{5.65}$$

Therefore, we want to prove (5.64) under the assumption that (5.65) holds for all $W, r \in \mathbb{N}$.

We use the notation

$$A(N) := \frac{1}{N}\sum_{n=1}^{N} \Lambda(n) F(\widetilde{g}_1^{\lfloor a_1(n)\rfloor} \cdot \ldots \cdot \widetilde{g}_k^{\lfloor a_k(n)\rfloor}\widetilde{x}),$$

and

$$B_{W,b}(N) := \frac{1}{N}\sum_{n=1}^{N} F(\widetilde{g}_1^{\lfloor a_1(Wn+b)\rfloor} \cdot \ldots \cdot \widetilde{g}_k^{\lfloor a_k(Wn+b)\rfloor}\widetilde{x})$$

for convenience.

Let $\varepsilon > 0$. Observe that the sequence $\psi(\mathbf{n}) = F(\widetilde{g}_1^{n_1} \cdot \ldots \cdot \widetilde{g}_k^{n_k}\widetilde{x})$ is an s-step nilsequence in k-variables. We apply Lemma 2.3.9 to deduce that there exists a system $(X', \mathcal{X}', \mu, S_1, \ldots, S_k)$ and functions $G_1, \ldots, G_s \in L^\infty(\mu)$ such that

$$\Big| F(\widetilde{g}_1^{n_1} \cdot \ldots \cdot \widetilde{g}_k^{n_k}\widetilde{x}) - \int \prod_{j=1}^{s} \Big(\prod_{i=1}^{k} S_i^{\ell_j n_i} \Big) G_j \, d\mu \Big| \leq \varepsilon$$

for all $n_1, \ldots, n_k \in \mathbb{Z}$, where $\ell_j = (s+1)!/j$.

Thus, if we define

$$A'(N) := \frac{1}{N}\sum_{n=1}^{N} \Lambda(n) \int \prod_{j=1}^{s+1} \Big(\prod_{i=1}^{k} S_i^{\ell_j \lfloor a_i(n)\rfloor} \Big) G_j \, d\mu,$$

and

$$B'_{W,b}(N) = \frac{1}{N}\sum_{n=1}^{N} \int \prod_{j=1}^{s+1} \Big(\prod_{i=1}^{k} S_i^{\ell_j \lfloor a_i(Wn+b)\rfloor} \Big) G_j \, d\mu,$$

we deduce that $|B_{W,b}(N) - B'_{W,b}(N)| \leq \varepsilon$, for all $N \in \mathbb{N}$, whereas $|A(N) - A'(N)| \leq \varepsilon(1 + o_N(1))$, by the prime number theorem.

The functions $a_1, \ldots, a_k$ satisfy the assumptions of Theorem 1.3.1. Thus, we deduce that if we pick w_0 (which provides a corresponding W_0) sufficiently large and apply the Cauchy-Schwarz inequality, we will get

$$\max_{\substack{1\leq b\leq W \\ (b,W_0)=1}} \Big| \frac{1}{N}\sum_{n=1}^{N} \big(\Lambda_{w_0,b}(n) - 1\big) \int \prod_{j=1}^{s+1} \Big(\prod_{i=1}^{k} S_i^{\ell_j \lfloor a_i(W_0 n+b)\rfloor} \Big) G_j \, d\mu \Big| \leq \varepsilon \tag{5.66}$$

for every sufficiently large $N \in \mathbb{N}$. In addition, we use (5.65), the inequality $|B_{W_0,b}(N) - B'_{W_0,b}(N)| \leq \varepsilon$ and the triangle inequality to infer that for N large enough, we have

$$\Big| B'_{W_0,b}(N) - \int_Y F \, dm_Y \Big| \leq 2\varepsilon, \tag{5.67}$$

for all $1 \leq b \leq W_0$ coprime to W_0.

Observe that (5.66) implies that for all N sufficiently large, we have

$$\Big| A'(W_0 N) - \frac{1}{\phi(W_0)} \sum_{\substack{1\leq b\leq W_0 \\ (b,W_0)=1}} B'_{W_0,b}(N) \Big| \leq 2\varepsilon,$$

and we can combine this with (5.67) to conclude that

$$\Big| A'(W_0 N) - \int_Y F \, dm_Y \Big| \leq 4\varepsilon$$

for all N sufficiently large. Since $|A'(N) - A(N)| \leq \varepsilon(1 + o_N(1))$, we finally arrive at the inequality

$$\Big| A(W_0 N) - \int_Y F \, dm_Y \Big| \leq 6\varepsilon,$$

for all large enough $N \in \mathbb{N}$. Since $|A(W_0 N) - A(W_0 N + b)| = o_N(1)$ for all $1 \le b \le W$, we conclude that

$$\left| A(N) - \int_Y F \, dm_Y \right| \le 7\varepsilon,$$

for all sufficiently large $N \in \mathbb{N}$. Sending $\varepsilon \to 0$, we deduce (5.64), which is what we wanted to show. □

Proof of Proposition Corollary 1.3.13. The result follows readily from Theorem 1.3.12. The first hypothesis of the criterion is satisfied, since each of the functions $a_i(t)$ satisfies (1.28), while condition (b) follows from Theorem 1.2.1 and our assumption that $a_i(Wt + b)$ belongs to $\mathcal{H}$. □

5.6 More general iterates

In this last section, we discuss how the hypotheses that the functions $a_i(t)$ in the iterates belong to a Hardy field $\mathcal{H}$ can be weakened. The starting point is Proposition 5.3.1, which was established for general smooth functions, subject to some growth inequalities on the derivative of some particular order (the integer k in the statement). Unfortunately, one cannot generalize theorems such as Theorem 1.3.4, which involve several functions to a more general class. The main obstruction is that in order to obtain the simultaneous Taylor expansions, one needs to find a function $L(t)$ (the length of the short interval) that satisfies a growth relation for all functions at the same time, which is non-trivial to perform, because we do not know how the derivatives of one function might grow relative to the derivatives of another function (this is where the assumption that all function belong to the same Hardy field is crucial). Nonetheless, this is still feasible in the case of one function, such as Theorem 1.3.3, which leads to Szemerédi-type results.

Proposition 5.6.1. *Let $a(t)$ be a function, defined for all sufficiently large t and satisfying $|a(t)| \to +\infty$, as $t \to +\infty$. Suppose there exists a positive integer k for which a is C^{k+1}, $a^{(k+1)}(t)$ converges to 0 monotonically, and such that*[10]

$$t^{5/8} \ll \left| a^{(k)}(t) \right|^{-\frac{1}{k}} \lll \left| a^{(k+1)}(t) \right|^{-\frac{1}{k+1}} \ll t.$$

Then, for any $\ell \in \mathbb{N}$, measure-preserving system $(X, \mathcal{X}, \mu, T_1, \dots, T_\ell)$, and functions $f_1, \dots, f_\ell \in L^\infty(\mu)$, we have

$$\lim_{w \to +\infty} \limsup_{N \to +\infty} \max_{\substack{1 \le b \le W \\ (b,W)=1}} \left\| \frac{1}{N} \sum_{n=1}^{N} (\Lambda_{w,b}(n) - 1) T_1^{\lfloor a(Wn+b) \rfloor} f_1 \cdot \ldots \cdot T_\ell^{\lfloor a(Wn+b) \rfloor} f_\ell \right\|_{L^2(\mu)} = 0.$$

We remark that any improvement in the parameter $5/8$ in Theorem K will also lower the term $t^{5/8}$ on the leftmost part of the growth inequalities accordingly.

Sketch of the proof of Proposition 5.6.1. We define $L(t)$ to be the geometric mean of the functions $\left| a^{(k)}(t) \right|^{-\frac{1}{k}}$ and $\left| a^{(k+1)}(t) \right|^{-\frac{1}{k+1}}$, which is well-defined for all t sufficiently large. A standard computation implies the relation

$$t^{5/8} \ll \left| a^{(k)}(t) \right|^{-\frac{1}{k}} \lll L(t) \lll \left| a^{(k+1)}(t) \right|^{-\frac{1}{k+1}} \ll t.$$

Regarding the parameter w as fixed, it suffices to show that

$$\limsup_{N \to +\infty} \max_{\substack{1 \le b \le W \\ (b,W)=1}} \left\| \frac{1}{N} \sum_{n=1}^{N} (\Lambda_{w,b}(n) - 1) T_1^{\lfloor g(Wn+b) \rfloor} f_1 \cdot \ldots \cdot T_\ell^{\lfloor g(Wn+b) \rfloor} f_\ell \right\|_{L^2(\mu)} = o_w(1).$$

[10] See the notational conventions for the notation $\lll$.

This follows if we show that

$$\limsup_{N\to+\infty} \max_{\substack{1\leq b\leq W\\ (b,W)=1}} \Big\| \mathop{\mathbb{E}}_{N\leq n\leq N+L(N)} \big(\Lambda_{w,b}(n)-1\big) T_1^{\lfloor a(Wn+b)\rfloor} f_1 \cdot \ldots \cdot T_\ell^{\lfloor a(Wn+b)\rfloor} f_\ell \Big\|_{L^2(\mu)} = o_w(1).$$

This derivation is very similar to the proof of [13, Lemma 4.3], which was stated only for bounded sequences. This is proven by covering the interval $[1, N]$ with non-overlapping sub-intervals that have the form $[m, m+L(m)]$ (for m large enough), where the term of the average on the last set of the covering is bounded as in (5.35). [11]

Using Proposition 5.3.1 and the abbreviated notation $g_{W,b}(t)$ for the function $g(Wt+b)$, we deduce that we can write

$$\lfloor g_{W,b}(n)\rfloor = \left\lfloor g_{W,b}(N) + \cdots + \frac{(n-N)^k g_{W,b}^{(k)}(N)}{k!} \right\rfloor$$

for all except at most $O(L(N)\log^{-100} N)$ values of $n \in [N, N+L(N)]$. Furthermore, we also have the equidistribution assumption of Proposition 5.3.1, which implies that Proposition 5.2.3 is applicable for the polynomial

$$g_{W,b}(N) + \cdots + \frac{(n-N)^k g_{W,b}^{(k)}(N)}{k!}$$

appearing in the iterates. The conclusion then follows similarly as in the proof of Theorem 1.3.1, so we omit the rest of the details. □

An application of the previous comparison is for the class of *tempered* functions, which we define promptly.

Definition 5.6.2. *Let i be a non-negative integer. A real-valued function g which is $(i+1)$-times continuously differentiable on (t_0,∞) for some $t_0 \geq 0$, is called a* tempered function of degree i *(we write $d_g = i$), if the following hold:*

(a) *$g^{(i+1)}(t)$ tends monotonically to 0 as $t\to\infty$;*

(b) $\lim_{t\to+\infty} t|g^{(i+1)}(t)| = +\infty$.

Tempered functions of degree 0 are called Fejér *functions.*

For example, consider the functions

$$g_1(t) = t^{1/25}(100+\sin\log t)^3,\ g_2(t) = t^{1/25},\ g_3(t) = t^{17/2}(2+\cos\sqrt{\log t}). \tag{5.68}$$

We have that g_1 and g_2 are Fejér, g_3 is tempered of degree 8 (which is not Hardy, see [2]). Every tempered function of degree i is eventually monotone and it grows at least as fast as $t^i \log t$ but slower than t^{i+1} (see [2]), so that, under the obvious modification of the definition of a strongly non-polynomial, tempered functions $\mathcal{T}$ are strongly non-polynomial. Also, for every tempered function g, we have that $(g(n))_{n\in\mathbb{N}}$ is equidistributed mod 1.[12]

In general, it is more restrictive to work with tempered functions than working with Hardy field ones. To see this, notice that ratios of tempered functions need not have limits, in contrast to the Hardy field case. For example, the functions g_1 and g_2 in (5.68) are such that $g_1(t)/g_2(t)$ has no limit as $t\to+\infty$. This issue persists even when we are dealing with a single function, as ratios that involve derivatives of the same function may not have a limit either. Indeed, we can easily see that g_1 from (5.68) (which was first studied in [11]) has the property that $\frac{tg_1'(t)}{g_1(t)}$ does not have a limit as $t\to+\infty$. The existence of the limit of the latter is important as it allows us to compare (via L' Hôpital's rule) growth rates of derivatives of functions with comparable growth rates.

[11] In particular, this case is much simpler than the method used to establish Theorem 1.3.1, in that we do not have to consider the more complicated double averaging scheme. In addition, we do not need any assumptions on $L(t)$ other than it is positive and $L(t) \prec t$.

[12] For Fejér functions this is a classical result due to Fejér (for a proof see [37]). The general case follows inductively by van der Corput's difference theorem.

In order to sidestep the aforementioned problematic cases, we restrict our study to the following subclass of tempered functions (see also [2], [35]).

Let $\mathcal{R} := \Big\{ g \in C^\infty(\mathbb{R}^+) :\ \lim_{t\to+\infty} \frac{tg^{(i+1)}(t)}{g^{(i)}(t)} \in \mathbb{R} \ \text{ for all } \ i \in \mathbb{N}\cup\{0\} \Big\}$;

$\mathcal{T}_i := \Big\{ g \in \mathcal{R} :\ \exists\, i < \alpha < i+1,\ \lim_{t\to+\infty} \frac{tg'(t)}{g(t)} = \alpha,\ \lim_{t\to+\infty} g^{(i+1)}(t) = 0 \Big\}$;

and $\mathcal{T} := \bigcup_{i=0}^{\infty} \mathcal{T}_i$. For example, $g_2 \in \mathcal{T}_0$ and $g_3 \in \mathcal{T}_8$ (g_2, g_3 are those from (5.68)).

Notice that while the class of Fejér functions contain sub-fractional functions, $\mathcal{T}_0$ does not as, according to [10, Lemma 6.4], if $g \in \mathcal{T}$ with $\lim_{t\to+\infty} \frac{tg'(t)}{g(t)} = \alpha$, then for every $0 < \beta < \alpha$ we have $t^\beta \prec g(t)$.

We will prove a convergence result for the class $\mathcal{T}$ through an application of Proposition 5.6.1.

Lemma 5.6.3. *Let g be a function in $\mathcal{T}$ and $0 < c < 1$. Then, for all large enough positive integers k, we have*

$$t^c \prec \left|g^{(k)}(t)\right|^{-\frac{1}{k}} \lll \left|g^{(k+1)}(t)\right|^{-\frac{1}{k+1}} \prec t.$$

Proof. Since $g(t) \prec t^{d_g+1}$ and $0 < c < 1$, we have $g(t) \prec t^{k(1-c)}$ for all large enough $k \in \mathbb{N}$, which implies

$$\frac{g^{(k)}(t)}{t^{-ck}} = \frac{g(t)}{t^{k(1-c)}} \cdot \prod_{i=1}^{k} \frac{tg^{(i)}(t)}{g^{(i-1)}(t)} \to 0.$$

Hence, $g^{(k)}(t) \prec t^{-ck}$ or, equivalently, $t^c \prec \left|g^{(k)}(t)\right|^{-\frac{1}{k}}$.

For the aforementioned k's, let $0 < q < 1$ so that $t^{kq} \prec g(t)$. Since $\lim_{t\to+\infty} \frac{tg'(t)}{g(t)} \notin \mathbb{N}$,

$$\frac{t^{k(q-1)}}{g^{(k)}(t)} = \frac{t^{kq}}{g(t)} \cdot \prod_{i=1}^{k} \frac{g^{(i-1)}(t)}{tg^{(i)}(t)} \to 0,$$

so $t^{k(q-1)} \prec g^{(k)}(t)$. As $\lim_{t\to+\infty} \frac{tg^{(k+1)}(t)}{g^{(k)}(t)} \in \mathbb{R}\setminus\{0\}$, we get $g^{(k+1)}(t) \ll t^{-1} g^{(k)}(t)$, so, if we let $\delta = \frac{q}{k+1}$, we have

$$\frac{\left|g^{(k+1)}(t)\right|^{-\frac{1}{k+1}}}{\left|g^{(k)}(t)\right|^{-\frac{1}{k}}} \gg \frac{t^{\frac{1}{k+1}}\left|g^{(k)}(t)\right|^{-\frac{1}{k+1}}}{\left|g^{(k)}(t)\right|^{-\frac{1}{k}}} = t^{\frac{1}{k+1}}\left|g^{(k)}(t)\right|^{\frac{1}{k(k+1)}} \succ t^{\frac{1}{k+1}} \cdot t^{\frac{q-1}{k+1}} = t^\delta,$$

completing the proof of the lemma (the rightmost inequality follows by [10]). □

Using Proposition 5.6.1 and [13, Theorem 2.2] we get the following result. More precisely, we use the fact here that [13, Theorem 2.2] holds for a single function a which has the property that, for some $k \in \mathbb{N}$, a is C^{k+1}, $a^{(k+1)}(t)$ converges to 0 monotonically, $1/t^k \prec a^{(k)}(t)$, and $|a^{(k)}(t)|^{-1/k} \prec |a^{(k+1)}(t)|^{-1/(k+1)}$ (see comments in [13, Subsection 2.1.5]). We omit its proof as it is identical to the one of Theorem 1.3.3.

Theorem 5.6.4. *Let $g \in \mathcal{T}$. For any $k \in \mathbb{N}$, measure-preserving system $(X, \mathcal{X}, \mu, T)$, and functions $f_1, \dots, f_k \in L^\infty(\mu)$, we have*

$$\lim_{N\to+\infty} \frac{1}{\pi(N)} \sum_{p\in\mathbb{P}:\ p\le N} T^{\lfloor g(p)\rfloor} f_1 \cdot \ldots \cdot T^{k\lfloor g(p)\rfloor} f_k = \lim_{N\to+\infty} \frac{1}{N} \sum_{n=1}^{N} T^n f_1 \cdot \ldots \cdot T^{kn} f_k, \tag{5.69}$$

where the convergence takes place in $L^2(\mu)$.

As in the Hardy field case, we have the corresponding recurrence result.

Theorem 5.6.5. *Let $g \in \mathcal{T}$. For any $k \in \mathbb{N}$, measure-preserving system $(X, \mathcal{X}, \mu, T)$, and set A with positive measure, we have*

$$\lim_{N\to+\infty} \frac{1}{\pi(N)} \sum_{p\in\mathbb{P}:\ p\le N} \mu(A \cap T^{-\lfloor g(p)\rfloor} A \cap \cdots \cap T^{-k\lfloor g(p)\rfloor} A) > 0.$$

The latter implies the following corollary, which guarantees arbitrarily long arithmetic progressions, with steps coming from the class of tempered functions evaluated at primes.

Corollary 5.6.6. *Let $g \in \mathcal{T}$. For any set $E \subseteq \mathbb{N}$ of positive upper density, and $k \in \mathbb{N}$, we have*

$$\liminf_{N \to +\infty} \frac{1}{\pi(N)} \sum_{p \in \mathbb{P}\colon\, p \leq N} \bar{d}\big(E \cap (E - \lfloor g(p) \rfloor) \cap \cdots \cap (E - k\lfloor g(p) \rfloor)\big) > 0.$$

Remark. In Theorem 5.6.4, and, thus, in Theorem 5.6.5 and Corollary 5.6.6, the floor function can be replaced with either the function $\lceil\cdot\rceil$ or the function $[[\cdot]]$. Furthermore, in each of these results, one can alternatively evaluate the sequences along the affine shifts $ap + b$, for $a, b \in \mathbb{R}$ with $a \neq 0$.

Bibliography

[1] V. Bergelson. Weakly mixing PET. *Ergodic Theory and Dynamical Systems*, **7**(3):337–349, 1987. 9, 21, 50

[2] V. Bergelson and I. Häland-Knutson. Weak mixing implies mixing of higher orders along tempered functions. *Ergodic Theory and Dynamical Systems*, **29**(5):1375–1416, 2009. 9, 138, 139

[3] V. Bergelson and A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *Journal of the American Mathematical Society*, **9**:725–753, 1996. 5

[4] V. Bergelson, J. Moreira, and F. K. Richter. Multiple ergodic averages along functions from a Hardy field: convergence, recurrence and combinatorial applications. Preprint 2020, arXiv:2006.03558. 8, 10

[5] M. Boshernitzan. Uniform distribution and Hardy fields. *Journal d'Analyse Mathématique*, **62**:225–240, 1994. 6, 7, 89

[6] M. Boshernitzan, R. L. Jones, and M. Wierdl. Integer and fractional parts of good averaging sequences in ergodic theory*. Convergence in Ergodic Theory and Probability.* Eds: V. Bergelson, P. March, J. Rosenblatt, by Walter de Gruyter & Co., Berlin, New York, 1996, 117-132. 21

[7] M. Boshernitzan, G. Kolesnik, A. Quas, and M. Wierdl. Ergodic averaging sequences. *Journal d'Analyse Mathématique*, **95**:63–103, 2005. 6

[8] Q. Chu. Multiple recurrence for two commuting transformations. *Ergodic Theory and Dynamical Systems*, **31**(3):771–792, 2011. 135

[9] L. Corwin and F. P. Greenleaf. *Representations of nilpotent Lie groups and their applications Part 1: Basic theory and examples.* Cambridge Studies in Advanced Mathematics. **18**, 1990. 30, 32

[10] S. Donoso, A. Koutsogiannis, and W. Sun. Joint ergodicity for functions of polynomial growth. To appear in *Israel Journal of Mathematics*. Preprint 2023, arXiv:2301.06911. 139

[11] S. Donoso, A. Koutsogiannis, and W. Sun. Pointwise multiple averages for sublinear functions. *Ergodic Theory and Dynamical Systems*, **40**:1594–1618, 2020. 138

[12] N. Frantzikinakis. Equidistribution of sparse sequences on nilmanifolds. *Journal d'Analyse Mathématique*, **109**(1):353–395, 2009. 11, 13, 20, 24, 90, 95, 104

[13] N. Frantzikinakis. Multiple recurrence and convergence for Hardy sequences of polynomial growth. *Journal d'Analyse Mathématique*, **112**(1):79–135, 2010. 7, 8, 9, 16, 18, 43, 76, 133, 138, 139

[14] N. Frantzikinakis. A multidimensional Szemerédi theorem for Hardy sequences of different growth. *Transactions of the American Mathematical Society*, **367**:5653–5692, 2012. 8, 17, 133

[15] N. Frantzikinakis. Some open problems on multiple ergodic averages. *Bulletin of the Hellenic Mathematical Society*, **60**:41–90, 2016. 13, 18

[16] N. Frantzikinakis. Joint ergodicity of fractional powers of primes. *Forum of Mathematics, Sigma*, **10**:e30, 2022. 15, 17, 20

[17] N. Frantzikinakis. Joint ergodicity of sequences. *Advances in Mathematics*, **417**:108918, 2023. 7, 8, 15, 30, 83, 89

[18] N. Frantzikinakis and B. Host. Weighted multiple ergodic averages and correlation sequences. *Ergodic Theory and Dynamical Systems*, **38**:81–142, 2015. 36

[19] N. Frantzikinakis, B. Host, and B. Kra. Multiple recurence and convergence for sequences related to the prime numbers. *Journal fur die Reine und Angewandte Mathematik*, **2007**, 2006. 14, 21

[20] N. Frantzikinakis, B. Host, and B. Kra. The polynomial multidimensional szemerédi theorem along shifted primes. *Israel Journal of Mathematics*, **194**:331–348, 2010. 14, 15, 21, 113

[21] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *Journal d'Analyse Mathématique*, **31**(1):204–256, 1977. 4

[22] H. Furstenberg and Y. Katznelson. An ergodic szemerédi theorem for commuting transformations. *Journal d'Analyse Mathématique*, 34:275–291, 1978. 5

[23] W. T. Gowers. A new proof of Szemerédi's theorem. *Geometric and Functional Analysis*, **11**(3):465–588, 2001. 28

[24] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics*, **167**(2):481–547, 2008. 14, 15

[25] B. Green and T. Tao. Linear equations in primes. *Annals of Mathematics*, **171**:1753–1850, 2010. 14

[26] B. Green and T. Tao. The Möbius function is strongly orthogonal to nilsequences. *Annals of Mathematics*, **175**(2):541–566, 2012. 14, 20

[27] B. Green and T. Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Annals of Mathematics*, **175**:465–540, 2012. 11, 21, 33, 34

[28] B. Green, T. Tao, and T. Ziegler. An inverse theorem for the Gowers $U^{s+1}[N]$-norm. *Annals of Mathematics*, **176**(2):1231–1372, 2012. 14, 36

[29] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Annals of Mathematics*, **161**(1):397–488, 2005. 5, 10, 27, 28

[30] B. Host and B. Kra. *Nilpotent structures in ergodic theory.* Mathematical Surveys and Monographs, Volume 236, American Mathematical Society, Providence, RI, 2018. 27, 28, 29, 30, 31, 33, 37

[31] M. N. Huxley. On the difference between consecutive primes. *Inventiones Mathematicae*, **15**:164–170, 1971. 37

[32] A.G. Khovanskii. *Fewnomials.* Translations of Mathematical Monographs, Volume 88, American Mathematical Society, 1991. 23

[33] A. Koutsogiannis. Closest integer polynomial multiple recurrence along shifted primes. *Ergodic Theory and Dynamical Systems*, **38**(2):666–685, 2018. 14, 18, 112, 114, 134

[34] A. Koutsogiannis. Integer part polynomial correlation sequences. *Ergodic Theory and Dynamical Systems*, **38**(4):1525–1542, 2018. 8

[35] A. Koutsogiannis. Multiple ergodic averages for tempered functions. *Discrete and Continuous Dynamical Systems*, **41**(3):1177–1205, 2021. 139

[36] A. Koutsogiannis and K. Tsinas. Ergodic averages for sparse sequences along primes. Preprint 2023, arXiv.2309.04939. 15, 16, 17, 18, 19, 20

[37] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences*. Pure and Applied Mathematics. Wiley-Interscience, New York-London-Sydney, 1974. 35, 120, 138

[38] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables. *Israel Journal of Mathematics*, **146**:303–315, 2005. 5, 10, 53, 64, 89

[39] A. Leibman. Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold. *Ergodic Theory and Dynamical Systems*, **25**(1):201–213, 2005. 10, 20, 31, 95

[40] A. Leibman. Orbits on a nilmanifold under the action of a polynomial sequence of translations. *Ergodic Theory and Dynamical Systems*, **27**(4):1239–1252, 2007. 10

[41] D. Leitmann. On the uniform distribution of some sequences. *Journal of the London Mathematical Society*, **14**:430–432, 1976. 15

[42] E. Lesigne. On the sequence of integer parts of a good sequence for the ergodic theorem. *Commentationes Mathematicae Universitatis Carolinae*, **36**(4):737–743, 1955. 21

[43] A. Mal'cev. On a class of homogeneous spaces. *Izvestiya Akad. Nauk SSSR, Ser Mat.*, **13**:9–32, 1949. 33

[44] K. Matomäki, X. Shao, T. Tao, and J. Teräväinen. Higher uniformity of arithmetic functions in short intervals i. all intervals. Preprint 2021, arXiv:2204.03754. 21, 37

[45] H. L. Montgomery and R. C. Vaughan. The large sieve. *Mathematika*, **20**:119–134, 1973. 38

[46] R. Nair. On polynomials in primes and J. Bourgain's circle method approach to ergodic theorems. *Ergodic Theory and Dynamical Systems*, **11**(3):485–499, 1991. 14

[47] W. Parry. Dynamical systems on nilmanifolds. Bulletin of the London Mathematical Society, **2**:37–40, 1970. 31

[48] M. Ratner. Raghunatan's topological conjecture and distribution of unipotent flows. *Duke Mathematical Journal*, **61**(1):235–280, 1991. 32

[49] F. Richter. Uniform distribution in nilmanifolds along functions from a Hardy field. *Journal d'Analyse Mathématique*, **149**:421–483, 2023. 11, 13, 72

[50] A. Sárközy. On difference sets of sequences of integers. III. *Acta Mathematica Hungarica*, **31**:355–386, 1978. 14

[51] I. Stux. On the uniform distribution of prime powers. *Communications on Pure and Applied Mathematics*, **27**:729–740, 1974. 15

[52] W. Sun. Multiple recurrence and convergence for certain averages along shifted primes. *Ergodic Theory and Dynamical Systems*, **35**(5):1592–1609, 2015. 14

[53] E. Szemerédi. On sets of integers containing k elements in arithmetic progression. *Acta Arithmetica*, **27**(1):199–245, 1975. 4

[54] K. Tsinas. Joint ergodicity of Hardy field sequences. *Transactions of the American Mathematical Society*, **376**:3191–3263, 2023. 8, 9, 10, 24

[55] K. Tsinas. Pointwise convergence in nilmanifolds along smooth functions of polynomial growth. *Ergodic Theory and Dynamical Systems*, Published online p:1-46. doi:10.1017/etds.2023.68, 2023. 12, 13

[56] M. Walsh. Norm convergence of nilpotent ergodic averages. *Annals of Mathematics*, **175**(3):1667–1688, 2012. 14

[57] M. Wierdl. Pointwise ergodic theorem along the prime number. *Israel Journal of Mathematics*, **64**:315–336, 1988. 14

[58] D. Wolke. Zur gleichverteilung einiger zahlenfolgen. *Mathematische Zeitschrift*, **142**:181–184, 1975. 15

[59] T. Wooley and T. Ziegler. Multiple recurrence and convergence along the primes. *American Journal of Mathematics*, **134**, 2010. 14

[60] T. Ziegler. Universal characteristic factors and furstenberg averages. *Journal of the American Mathematical Society*, 20:53–97, 2004. 5