ERGODIC AVERAGES FOR SPARSE SEQUENCES ALONG PRIMES

ANDREAS KOUTSOGIANNIS AND KONSTANTINOS TSINAS

Abstract. We investigate the limiting behavior of multiple ergodic averages along sparse sequences evaluated at prime numbers. Our sequences arise from smooth and well-behaved functions that have polynomial growth. Central to this topic is a comparison result between standard Cesáro averages along positive integers and averages weighted by the (modified) von Mangoldt function. The main ingredients are a recent result of Matomäki, Shao, Tao and Teräväinen on the Gowers uniformity of the latter function in short intervals, a lifting argument that allows one to pass from actions of integers to flows, a simultaneous (variable) polynomial approximation in appropriate short intervals, and some quantitative equidistribution results for the former polynomials. We derive numerous applications in multiple recurrence, additive combinatorics, and equidistribution in nilmanifolds along primes. In particular, we deduce that any set of positive density contains arithmetic progressions with step $\lfloor p^c \rfloor$, where c is a positive non-integer and p denotes a prime, establishing a conjecture of Frantzikinakis.

Contents

1. Introduction and main results

1.1. **Motivation.** The seminal work of Furstenberg [18] towards Szemerédi's theorem has initiated significant interest in the study of ergodic theoretic problems and their applications in problems of combinatorial or number-theoretic nature. Dynamical methods have proven extremely effective at tackling problems relating to the combinatorial richness of positive density subsets of integers. Furthermore, there are frequently no alternative methods that can recover the results that ergodic theoretic tools provide. For instance, we have far-reaching generalizations of Szemerédi's theorem, such as the Bergelson-Leibman theorem [2] that produces polynomial progressions on sets of integers with positive density.

The general structure of our problems is the following: we are given a collection of sequences $a_1(n), \ldots, a_k(n)$ of integers and a standard probability space $(X, \mathcal{X}, \mu)$ equipped with invertible, commuting, measure-preserving transformations $T_1, \ldots, T_k$ that act on

2020 *Mathematics Subject Classification.* Primary: 37A44; Secondary: 28D05, 11B30.

Key words and phrases. Ergodic averages, recurrence, prime numbers, Hardy fields.

The second author was supported by ELIDEK–Fellowship number 5367 (3rd Call for HFRI Ph.D. Fellowships).

X, and we examine the limiting behavior of the multiple averages

$$\frac{1}{N}\sum_{n=1}^{N} T_1^{a_1(n)}f_1\cdot\ldots\cdot T_k^{a_k(n)}f_k. \tag{1}$$

Throughout the article, these assumptions on the transformations will be implicit; we call the tuple $(X,\mathcal{X},\mu,T_1,\ldots,T_k)$ a *measure-preserving system* (or just *system*). Here $f_1,\ldots,f_k$ are functions in $L^\infty(\mu)$ and we concern ourselves with their convergence mainly in the L^2-sense. In view of Furstenberg's correspondence principle, a satisfactory answer to this problem typically ensures that sets with positive density possess patterns of the form $(m,m+a_1(n),\ldots,m+a_k(n))$, where $m,n\in\mathbb{N}$. Specializing to the case where all the sequences are equal and $T_i=T^i$, we arrive at the averages

$$\frac{1}{N}\sum_{n=1}^{N} T^{a(n)}f_1\cdot T^{2a(n)}f_2\cdot\ldots\cdot T^{ka(n)}f_k, \tag{2}$$

which relate to patterns of arithmetic progressions, whose common difference belongs to the set $\{a(n)\colon n\in\mathbb{N}\}$.

Furthermore, it is particularly tempting to conjecture that results pertaining to mean convergence of the averages in (1) should still be valid, if we restrict the range of summation to a sparse set such as the primes. Normalizing appropriately, we contemplate whether or not the averages

$$\frac{1}{\pi(N)}\sum_{p\in\mathbb{P}\colon p\leq N} T_1^{a_1(p)}f_1\cdot\ldots\cdot T_k^{a_k(p)}f_k \tag{3}$$

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.

The first results in this direction were established in the case $k=1$. Namely, Sárközy [40] used methods from analytic number theory to show that sets of positive density contain patterns of the form $(m,m+p-1)$, where p is a prime.[1] Additionally, Wierdl [45] established the even stronger pointwise convergence result for the averages (3) in the case $k=1$ and $a_1(n)=n$, while Nair generalized this theorem to polynomials evaluated at primes [38].

In the setting of several iterates, the first results were provided by Frantzikinakis, Host, and Kra [16], who established that sets of positive density contain 3-term arithmetic progressions whose common difference is a shifted prime. Furthermore, they demonstrated that the averages in (3) converge in the case $k=2$, $T_1=T_2$ and $a_i(n)=in,\ i\in\{1,2\}$. This was generalized significantly by Wooley and Ziegler [47] to hold in the case that the sequences $a_i(n),\ i\in\{1,\ldots,k\}$ are polynomials with integer coefficients and the transformations $T_1,\ldots,T_k$ are the same. Following that, Frantzikinakis, Host, and Kra confirmed the validity of the Bergelson-Leibman theorem in [17] along the shifted primes. In addition, they showed that the averages in (3) converge in norm when $a_i(n)$ are integer polynomials. Furthermore, Sun obtained convergence and recurrence results in [42] for a single transformation and iterates of the form $i\lfloor an\rfloor,\ i\in\{1,\ldots,k\}$ or $\lfloor jan\rfloor,\ j\in\{1,\ldots,k\}$, with a irrational. Finally, using the convergence results in [30] along $\mathbb{N}$ for integer parts of real polynomials and several transformations, the first author extended the convergence result of [17] to real polynomials in [29], obtaining recurrence for polynomials with real coefficients rounded to the closest integer. In all of the previous cases, combinatorial applications along the shifted primes were derived as well.

[1] Throughout this article, it will be a reoccurring theme that in combinatorial applications, certain arithmetic obstructions force one to consider the set of shifted primes $\mathbb{P}-1$ (or $\mathbb{P}+1$) in place of $\mathbb{P}$, when dealing with polynomials. This is a necessary assumption, as in such cases the corresponding results for the set $\mathbb{P}$ are easily seen to be incorrect (see, for example, [42, Remark 1.4]).

In the case of multiple iterates, a shared theme in the methods used has been the close reliance on the deep results provided by the work of Green and Tao in their effort to show that primes contain arbitrarily long arithmetic progressions [20]. For instance, all results[2] relied on the Gowers uniformity of the (modified) von Mangoldt function that was established in [21] conditional to two deep conjectures, which were subsequently verified in [24] and [22].

It was conjectured by Frantzikinakis that the polynomial theorems along primes should hold for more general sequences involving fractional powers n^c, such as $\lfloor n^{3/2} \rfloor$, $\left\lfloor n^{\sqrt{2}} \right\rfloor$ or even linear combinations thereof. Indeed, it was conjectured in [10] that the sequence $\lfloor p_n^c \rfloor$, where c is a positive non-integer and p_n is the sequence of primes is good for multiple recurrence and convergence. To be more precise, he conjectured that the averages

$$\frac{1}{\pi(N)} \sum_{p \in \mathbb{P}\colon p \leq N} T^{\lfloor p^c \rfloor} f_1 \cdot \ldots \cdot T^{k \lfloor p^c \rfloor} f_k \tag{4}$$

converge in $L^2(\mu)$ for all positive integers k and all positive non-integers c. Analogously, we have the associated multiple recurrence conjecture, namely that all sets of positive upper density contain k-term arithmetic progressions with common difference of the form $\lfloor p^c \rfloor$. When $0 < c < 1$, one can leverage the fact that the range of $\lfloor p_n^c \rfloor$ contains all sufficiently large integers to establish the multiple recurrence result. 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 [41] or [46] when $c < 1$ and [34] 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 [17] (concerning polynomials), but the Gowers uniformity of the von Mangoldt functions is insufficient to establish this claim. The other approach would be to use the method of characteristic factors, which is based on the structure theorem of Host-Kra [26]. Informally, this reduces the task of proving convergence to a specific class of systems with special algebraic structure called 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 more general averages of the form

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

for distinct positive non-integer $c_1, \ldots, c_k$. The recent result of Frantzikinakis [13] verifies 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, even in the more general case where the sequences in the iterates are linearly independent fractional polynomials. The number theoretic input required is a sieve-theoretic upper bound for the number of tuples of primes of a specific form, as well as an equidistribution result on fractional powers of primes in the torus that was already known. These methods relied heavily on the use of the joint ergodicity results in [14] and, thus, the linear independence assumption on the fractional polynomials was absolutely essential. In the same paper, it was conjectured [13, Problem] that the case of fractional polynomials can be generalized to a significantly larger class of functions of polynomial growth, called Hardy field functions, which we consider below. The conjecture asks for necessary and sufficient conditions so that the averages along primes converge to the product of the integrals in ergodic systems. The arguments in

[2] The methods in [47] do not invoke the full power of this theorem, although their approach draws heavily from the work of Green and Tao.

[13] cannot cover this larger class of functions,[3] as it was remarked in Subsection 1.3 of that article.

In this article, our objective is to strengthen the convergence results in [17] and [13] and resolve the convergence problem of the averages in (4). Actually, there is no advantage in confining ourselves to sequences of the form $\lfloor p^c \rfloor$, so we consider the more general class of sequences arising from Hardy field functions of polynomial growth (see Section 2 for the general definition), which, loosely speaking, are functions with pleasant behavior (such as smoothness, for instance). The prototypical example of a Hardy field is the field $\mathcal{LE}$ of logarithmico-exponential functions, which are defined by a finite combination of the operations $+, -, \times, \div$ and the functions $\exp, \log$ acting on a real variable t and real constants. For instance, the field $\mathcal{LE}$ contains the functions $\log^{3/2} t$, t^π, $t^{17}\log t + \exp(\sqrt{t^{\log t} + \log\log t})$. The fact that $\mathcal{LE}$ is a Hardy field was established in [25] and the reader can keep this in mind as a model case throughout this article. We resolve several conjectures involving the convergence of the averages in (3) along Hardy sequences. Consequently, we derive several applications in recurrence and combinatorics that expand the known results in the literature. Finally, we also establish an equidistribution result in nilmanifolds for sequences evaluated at primes.

1.2. **Main results.** We present here our main theorems. We start by stating our mean convergence results, followed by their applications to multiple recurrence and combinatorics, and conclude our presentation with the equidistribution results in nilmanifolds. We will assume below that we are working with a Hardy field $\mathcal{H}$ that contains the polynomial functions. This assumption is not necessary, but it simplifies the proofs of our main theorems. Besides, this restriction is very mild and the most interesting Hardy fields contain the polynomials. 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 $\mathbb{N}$ in the literature that were proved under these hypotheses and we will not need to use the implied additional structure on $\mathcal{H}$ in any of our arguments.

1.2.1. *Comparison between averaging schemes.* For many number-theoretic problems, a suitable proxy for capturing the distribution of the prime numbers is the von-Mangoldt function, which is defined on $\mathbb{N}$ by

$$\Lambda(n) = \begin{cases} \log p & , \text{ if } n = p^k \text{ for some prime } p \text{ and } k \in \mathbb{N} \\ 0 & , \text{ otherwise} \end{cases}. \tag{5}$$

The function Λ has mean value 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 one can think of Λ as being 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

$$\Lambda_{w,b}(n) = \frac{\phi(W)}{W} \Lambda(Wn + b), \tag{6}$$

where ϕ denotes the Euler totient function.

Our main result provides a comparison between ergodic averages along primes and averages along natural numbers. This will allow us to transfer mean convergence results for Cesàro averages to the prime setting, answering numerous conjectures regarding norm convergence of averages as those in (3) followed by applications in multiple recurrence

[3] A more fundamental obstruction in this more general setting was that the necessary seminorm estimates were unavailable even in the simplest case of averages along $\mathbb{N}$, apart from some known special cases. This was established a few months later by the second author [44].

and combinatorics. We explain the choice of the conditions on the functions a_{ij} in Subsection 1.3. Roughly speaking, the first condition implies that the sequence a_{ij} is equidistributed mod 1 due to a theorem of Boshernitzan (see Theorem D in Section 2).

Theorem 1.1. *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*

$$\lim_{t\to+\infty} \left| \frac{a_{ij}(t) - q(t)}{\log t} \right| = +\infty \quad \textit{for every polynomial } q(t) \in \mathbb{Q}[t], \tag{7}$$

or

$$\lim_{t\to+\infty} |a_{ij}(t) - q(t)| = 0 \quad \textit{for some polynomial } q(t) \in \mathbb{Q}[t] + \mathbb{R}. \tag{8}$$

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

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

Remark 1. *We can easily verify that each of the integer parts can be individually replaced by other rounding functions, such as the ceiling function (which we denote by $\lceil \cdot \rceil$) or the closest integer function (denoted by $[[\cdot]]$). This is an immediate consequence of the identities $\lceil x \rceil = -\lfloor -x \rfloor$ and $[[x]] = \lfloor x + 1/2 \rfloor$, for all $x \in \mathbb{R}$ and the fact that the affine shifts (by rationals) $q_1 a_{ij} + q_2, q_1, q_2 \in \mathbb{Q}$, still satisfy (7) or (8) if a_{ij} does.*

Theorem 1.1 is the main tool that we use to derive all of our applications. The bulk of the article is aimed towards establishing it and everything else is practically a corollary (in combination with known norm convergence theorems for Cesàro averages). We remark that unlike several of the theorems below, there are no "independence" assumptions between the functions a_{ij}, although, in applications, we will need to impose analogous assumptions to ensure convergence of the averages, firstly along $\mathbb{N}$, and then along $\mathbb{P}$. In order to clarify how the comparison works, we present the following theorem, which is effectively a corollary of Theorem 1.1 and which shall be proven in Section 6.

Theorem 1.2. *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 $1 \leq i \leq k,\ 1 \leq j \leq \ell$, $a_{ij} \in \mathcal{H}$ are functions of polynomial growth such that the following conditions are satisfied:*
(a) Each one of the functions $a_{ij}(t)$ satisfies either (7) or (8).
(b) For all positive integers W, b, the averages

$$\frac{1}{N} \sum_{n=1}^{N} \big(\prod_{i=1}^{k} T_i^{\lfloor a_{i1}(Wn+b) \rfloor} \big) f_1 \cdot \ldots \cdot \big(\prod_{i=1}^{k} T_i^{\lfloor a_{i\ell}(Wn+b) \rfloor} \big) f_\ell \tag{9}$$

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

Then, the averages

$$\frac{1}{\pi(N)} \sum_{p\in\mathbb{P}:\, p\leq N} \big(\prod_{i=1}^{k} T_i^{\lfloor a_{i1}(p) \rfloor} \big) f_1 \cdot \ldots \cdot \big(\prod_{i=1}^{k} T_i^{\lfloor a_{i\ell}(p) \rfloor} \big) f_\ell \tag{10}$$

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

Furthermore, if the averages in (9) 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 (10) is equal to F.

In the setting of Hardy field functions, 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 (7)) and it is straightforward to check that the function $a_{ij}(Wt + b)$ satisfies a similar growth condition as well. Therefore, one can think of the second condition morally as asking to establish convergence in the case $W = 1$.

The final part of Theorem 1.2 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 (although one can get a simplified expression in special cases, or under some total ergodicity assumptions on the system).

1.2.2. *Convergence of ergodic averages along primes.* The foremost application is that the averages in (2) converge when $a(n)$ is a Hardy sequence and when we average along primes. This will also lead to generalizations of Szemerédi's theorem in our applications. The following theorem is a corollary of our comparison and the convergence results in [10] (specifically, Theorems 2.1 and 2.2 of that paper). In conjunction with the corresponding recurrence result of Theorem 1.6 below, we get an affirmative answer to a stronger version of [10, Problem 7] (this problem also reappeared in [12, Problem 27]), which was stated only for sequences of the form $n^c, c \in \mathbb{R}^+ \setminus \mathbb{N}$.

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

$$\lim_{t\to+\infty} \left| \frac{a(t) - cq(t)}{\log t} \right| = +\infty \text{ for every } c \in \mathbb{R} \text{ and every } q \in \mathbb{Z}[t], \tag{11}$$

or

$$\lim_{t\to+\infty} |a(t) - cq(t)| = d \text{ for some } c, d \in \mathbb{R} \text{ and some } q \in \mathbb{Z}[t]. \tag{12}$$

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

$$\frac{1}{\pi(N)} \sum_{p\in\mathbb{P}:\ p\leq N} T^{\lfloor a(p)\rfloor} f_1 \cdot \ldots \cdot T^{k\lfloor a(p)\rfloor} f_k \tag{13}$$

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

In particular, if a satisfies (11), the limit of the averages in (13) is equal to the limit in $L^2(\mu)$ of the averages

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

Comment. We can replace the floor function in (13) with either the function $\lceil\cdot\rceil$ or the function $[[\cdot]]$. The assumption that the iterates are Hardy field functions can also be relaxed. We discuss this more in Section 7.

Observe that there is only one function appearing in the statement of the previous theorem. The following convergence results concern the case where we may have several different Hardy field functions. In both cases, there are some "independence" assumptions between the functions involved, which has the advantage of providing an exact description of the limit for the averages along $\mathbb{N}$. Thus, we can get a description for the limit along $\mathbb{P}$ 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. Theorem 1.1 combines with [44, Theorem 1.2] to provide the next result. This generalizes the theorem of Frantzikinakis [13, Theorem 1.1] and gives a positive answer to [13, Problem]. Unlike the previous theorem, we have to impose here an additional assumption on $\mathcal{H}$, since the respective convergence result along $\mathbb{N}$ is established under this condition. The field $\mathcal{LE}$ does not have the property appearing in the ensuing theorem, but it is contained in the Hardy field of Pfaffian functions, which does (for the definition, see [44, Section 2]).

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.*[4] *For a positive integer k, let $a_1, \dots, a_k$ be functions of polynomial growth and assume that every non-trivial linear combination a of them satisfies*

$$\lim_{t\to+\infty}\Big|\frac{a(t)-q(t)}{\log t}\Big| = +\infty \text{ for every } q(t)\in\mathbb{Z}[t].^5 \tag{14}$$

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

$$\lim_{N\to+\infty}\frac{1}{\pi(N)}\sum_{p\in\mathbb{P}:\, p\leq N} T^{\lfloor a_1(p)\rfloor}f_1\cdot\ldots\cdot T^{\lfloor a_k(p)\rfloor}f_k = \tilde{f}_1\cdot\ldots\cdot\tilde{f}_k, \tag{15}$$

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 $L^2(\mu)$.

Remark 2. *We remark that we can also transfer the convergence result appearing in* [43, Theorem 1.3] *to primes, although we do not have useful information on the limiting behavior of the associated averages to deduce recurrence results.*

In the case of several commuting transformations, knowledge of the limiting behavior for averages along $\mathbb{N}$ is sparse. This is naturally a barrier to proving multidimensional analogs of our recurrence results below along primes. Nonetheless, we have the following convergence theorem, which adapts the convergence result in [11, Theorem 2.3] 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 1.5. *Let $k\in\mathbb{N}$, $\mathcal{H}$ be a shift-invariant Hardy field, $a_1, \dots, a_k$ be functions in $\mathcal{H}$ with pairwise distinct growth rates and such that there exist integers $d_i \geq 0$ satisfying*

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

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

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

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 1.4, we note that Theorem 1.5 shows that we have norm convergence in the case when $a_i(t) = t^{c_i}$ for distinct, positive non-integers c_i.

Comment. In the previous two theorems, we can replace the integer part with any other rounding function in each iterate individually (see Remark 1).

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

[5] Actually, we can assume here the more general condition that

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

for every $(t_1, \dots, t_k) \in [0,1)^k \setminus \{(0,\dots,0)\}$, where $e(x) = e^{2\pi i x}$, $x\in\mathbb{R}$ (see the remark under [44, Theorem 1.2]). This condition is necessary and sufficient in order for (15) to hold.

1.2.3. *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 to extrapolate 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 [29] 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. [10, Footnote 4]).

Our first application relates to the averages of the form as in (3). We have the following theorem.

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 (11)*, 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 (12) *with $cp(0)+d=0$,*[6] *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 $\mathbb{P}+1$.

We recall that for a subset E of $\mathbb{N}$, its upper density $\bar{d}(E)$ is defined by

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

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 (11)*, 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 (12) *with $cp(0)+d=0$, then the set*

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

has non-empty intersection with the sets $\mathbb{P}-1$ or $\mathbb{P}+1$.[7]

Specializing to the case where $a(n)=n^c$ where c is a positive non-integer, Theorem 1.3 and part (a) of Theorem 1.6 provide an affirmative answer to [12, Problem 27].

Remark 3. *In part (a) of both Theorem 1.6 and Corollary 1.7, one can evaluate the sequences along $p+u$ instead of p, for any $u\in\mathbb{Z}$, or even more generally 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* (11) *as well. However, the shifts $p-1$ and $p+1$ are the only correct ones in part (b) of Theorem 1.6. Notice also that the function $\lfloor\cdot\rfloor$ can be replaced by $\lceil\cdot\rceil$ or $[[\cdot]]$ in part (a) of the two previous statements.*

Now, we state the recurrence result obtained by Theorem 1.4.

[6] Notice here the usual necessary assumption that we have to postulate on the polynomial, i.e., to have no constant term, in order to obtain a recurrence, and, consequently, a combinatorial result.

[7] In this case only, $\bar{d}(E)$ can be replaced by $d^*(E) := \limsup_{|I|\to+\infty} \frac{|E\cap I|}{|I|}$ following the arguments from [29], where the lim sup is taken along all intervals $I\subseteq\mathbb{Z}$ with lengths tending to infinity.

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 $a_1, \ldots, a_k \in \mathcal{H}$ are functions of polynomial growth whose non-trivial linear combinations satisfy* (14). *Then, for any measure-preserving system $(X, \mathcal{X}, \mu, T)$, and set A with positive measure, we have that*

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

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 1.8, we have*

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

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 1.8 and relies on the convergence result of Theorem 1.5.

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 1.5. Then, for any system $(X, \mathcal{X}, \mu, T_1, \ldots, T_k)$ and set A with positive measure, we have that*

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

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

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

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 1.10 and vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{Z}^d$, we have*

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

Comment. Once again, we remark that in the recurrence results in both Theorem 1.8 and Theorem 1.10 and the corresponding corollaries, one can replace p with any other affine shift $ap + b$ with $a, b \in \mathbb{Q}$ $(a \neq 0)$, as we explained in Remark 3. In addition, one can replace the floor functions with either $\lceil\cdot\rceil$ or $[[\cdot]]$.

1.2.4. *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 is similar in spirit to Theorem 1.2.

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 $1 \leq i \leq k$, the function $a_i(t)$ satisfies either (7) *or* (8).
(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 [22, 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 [33]). Nonetheless, in this case, we do not have a nice description for the orbit of this polynomial sequence.

On the other hand, equidistribution results in higher-step nilmanifolds (along primes) for sequences such as $\lfloor n^c\rfloor$, with c a non-integer ($c > 1$), are unknown even in the simplest case of one fractional power. Theorem 1.12 will allow us to obtain the first results in this direction from the corresponding results along $\mathbb{N}$. Equidistribution results for Hardy sequences along $\mathbb{N}$ were obtained originally by Frantzikinakis in [9], while more recently new results were established by Richter [39] and the second author [43]. In view of the structure theory of Host-Kra [26], results of this nature are essential to demonstrate that the corresponding multiple ergodic averages along $\mathbb{N}$ converge in $L^2(\mu)$. All of the pointwise convergence theorems that we mentioned above can be transferred to the prime setting. As an application, we state the following sample corollary of Theorem 1.12. 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 1.13. *Let k be a positive integer, $\mathcal{H}$ be a Hardy field invariant under affine shifts, and suppose that $a_1, \ldots, a_k \in \mathcal{H}$ are functions of polynomial growth, for which there exists an $\varepsilon > 0$, so that every non-trivial linear combination a of them satisfies*

$$\lim_{t\to+\infty}\Big|\frac{a(t)-q(t)}{t^{\varepsilon}}\Big| = +\infty \text{ for every } q(t) \in \mathbb{Z}[t]. \tag{16}$$

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

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

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

The assumption in (16) is a byproduct of the corresponding equidistribution result along $\mathbb{N}$ proven in [43]. Also, the assumption on $\mathcal{H}$ can be dropped since the arguments in [43] rely on some growth assumptions on the functions a_i which translate over to their shifted versions. We choose not to remove the assumption here since the results in [43] are not stated in this setup.

Our corollary implies that the sequence

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

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 $c_1, \ldots, c_k$ and for all points $x_i \in X_i$. This is stronger than the result of Frantzikinakis [13] 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 (16) 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 [9].

1.3. **Strategy of the proof and organization.** The bulk of the paper is spent on establishing the asserted comparison between the W-tricked averages and the standard Cesàro averages (Theorem 1.1). The main trick is to recast our problem to the setting where our averages range over a short interval of the form $[N, N + L(N)]$, where $L(t)$ is a function of sub-linear growth chosen so that Hardy sequences are approximated sufficiently well by polynomials in these intervals. Naturally, the study of the primes in short intervals requires strong number theoretic input and this is provided by the recent result in [36] on the Gowers uniformity of several arithmetic functions in short intervals (this is Theorem A in the following section). The strategy of restricting ergodic averages to short intervals was first used by Frantzikinakis in [10] to demonstrate the convergence of the averages in (2) when $a(n)$ is a Hardy sequence and then amplified further by the second author in [44] to resolve the problem in the more general setting of the averages in (1) (for one transformation). Certainly, the uniformity estimate in Theorem A requires that the interval is not too short, but it was observed in [44] that one can take the function $L(t)$ to grow sufficiently fast, as long as one is willing to tolerate polynomial approximations with much larger degrees.

After this step has been completed, one typically employs a complexity reduction argument (commonly referred to as PET induction in the literature) that relies on repeated applications of the van der Corput inequality. Using this approach, one derives iterates that are comprised of several expressions with integer parts, which are then assembled together using known identities for the floor function (with an appropriate error). This approach was used for the conventional averages over $\mathbb{N}$ in [10] and [44], because one can sloppily combine integer parts in the iterates at the cost of inserting a bounded weight in the corresponding averages. To be more precise, this weight is actually the characteristic function of a subset of $\mathbb{N}$. However, we cannot afford to do this blindly in our setting, since there is no guarantee that this subset of $\mathbb{N}$ does not correlate very strongly with $\Lambda_{w,b}(n) - 1$, which could imply that the resulting average is large. The fact that the weight $\Lambda_{w,b} - 1$ is unbounded complicates this step as well. Nonetheless, it was observed in [29] (using an argument from [30]),[8] that if the fractional parts of the sequences in the iterates do not concentrate heavily around 1, then one can 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. Since there are no nuisances with combining rounding functions in the iterates, one can then run the complexity reduction argument in the new system and obtain the desired bounds.

Unfortunately, there is still an obstruction in this approach arising from the fact that the flows in the extension are not continuous. To be more precise, let us assume that we derived an approximation of the form $a(n) = p_N(n) + \varepsilon_N(n)$, where $n \in [N, N + L(N)]$, $p_N(n)$ is a Taylor polynomial and $\varepsilon_N(n)$ is the remainder term. The PET induction can eliminate the polynomials $p_N(n)$, by virtue of the simple observation that taking sufficiently many "discrete derivatives" makes a polynomial vanish. However, this procedure cannot eliminate the error term $\varepsilon_N(n)$ at all and the fact that the flow is not continuous prohibits us from replacing them with zero. Thus, we take action to discard $\varepsilon_N(n)$ beforehand. This is done by studying the equidistribution properties of the polynomial $p_N(n)$ in the prior approximation, using standard results from the equidistribution theory of finite polynomial orbits due to Weyl. Practically, we show that for "almost all" values of n in the interval $[N, N + L(N)]$, we can write $\lfloor p_N(n) + \varepsilon_N(n) \rfloor = \lfloor p_N(n) \rfloor$, so that the error $\varepsilon_N(n)$ can be removed from the expressions in the iterates.

[8] This argument was first used for $k = 1$ in [5] and [35] to prove that when a sequence of real positive numbers is good for (single term) pointwise convergence, then its floor value is also good. The method was later adapted to the $k = 2$ setting by Wierdl (personal communication with the first author, 2015).

In our approach, some equidistribution assumptions on our original functions are required. This clarifies the conditions on Theorem 1.1. Indeed, (7) implies that the sequence $\big(a_{ij}(n)\big)_n$ is equidistributed modulo 1 (due to Theorem D), while condition (8) implies that the function $a_{ij}(t)$ is essentially equal to a polynomial with rational coefficients (thus periodic modulo 1).

1.3.1. *A simple example.* We demonstrate the methods discussed above in a basic case that avoids most complications that appear in the general setting. Even this simple case, however, is not covered by prior methods in the literature. We will use some prerequisites from the following section, such as Theorem A.

We consider the averages

$$\frac{1}{\pi(N)}\sum_{p\in\mathbb{P}\colon p\leq N} T^{\lfloor p^{3/2}\rfloor}f_1\cdot T^{2\lfloor p^{3/2}\rfloor}f_2, \tag{17}$$

where $(X,\mathcal{X},\mu,T)$ is a system and $f_1,f_2\in L^\infty(\mu)$. For every $1\leq b\leq W$ with $(b,W)=1$, we study the averages

$$\frac{1}{N}\sum_{n=1}^{N}\big(\Lambda_{w,b}(n)-1\big)T^{\lfloor n^{3/2}\rfloor}f_1\cdot T^{2\lfloor n^{3/2}\rfloor}f_2, \tag{18}$$

which is the required comparison for the averages in (17). We will show that as $N\to+\infty$ and then $w\to+\infty$, the norm of this average converges to 0 uniformly in b.

We set $L(t)=t^{0.65}$. Notice that $L(t)$ grows faster than $t^{5/8}$, which is a necessary condition to use Theorem A. In order to establish the required convergence for the averages in (18), it suffices to show that

$$\limsup_{r\to+\infty}\Big\lVert \mathop{\mathbb{E}}_{r\leq n\leq r+L(r)}\big(\Lambda_{w,b}(n)-1\big)T^{\lfloor n^{3/2}\rfloor}f_1\cdot T^{2\lfloor n^{3/2}\rfloor}f_2\Big\rVert_{L^2(\mu)}=o_w(1) \tag{19}$$

uniformly in b. We remark that in the more general case that encompasses several functions, we will need to average over the parameter r as well and, thus, we are dealing with a double-averaging scheme. This reduction is the content of Lemma 5.1.

Using the Taylor expansion around r, we can write for every $0\leq h\leq L(r)$:

$$(r+h)^{3/2}=r^{3/2}+\frac{3r^{1/2}h}{2}+\frac{3h^2}{8r^{1/2}}-\frac{3h^3}{48\xi_h^{3/2}},\quad\text{where } \xi_h\in[r,r+h].$$

Observe that the error term is smaller than a constant multiple of

$$\frac{\big(L(r)\big)^3}{r^{3/2}}=o_r(1).$$

We show that we have

$$\Big\lfloor (r+h)^{3/2}\Big\rfloor=\Bigg\lfloor r^{3/2}+\frac{3r^{1/2}h}{2}+\frac{3h^2}{8r^{1/2}}\Bigg\rfloor$$

for "almost all" $0\leq h\leq L(r)$, in the sense that the number of h's that do not obey this relation is bounded by a constant multiple of $L(r)\log^{-100} r$ (say). Thus, their contribution on the average is negligible, since the sequence $\Lambda_{w,b}$ has size comparable to $\log r$.

Let us denote by $p_r(h)$ the quadratic polynomial in the Taylor expansion above. In order to establish this assertion, we will investigate the discrepancy of the finite sequence $\big(p_r(h)\big)_{0\leq h\leq L(r)}$, using some exponential sum estimates and the Erdős-Turán inequality (Theorem E). This is the content of Proposition 4.4.

Assuming that all the previous steps were completed, we shall ultimately reduce our problem to showing that

$$\limsup_{r\to+\infty}\Big\|\mathop{\mathbb{E}}_{0\leq h\leq L(r)}\big(\Lambda_{w,b}(r+h)-1\big)T^{\lfloor p_r(h)\rfloor}f_1\cdot T^{2\lfloor p_r(h)\rfloor}f_2\Big\|_{L^2(\mu)}=o_w(1)$$

uniformly for $1\leq b\leq W$ coprime to W. Now that the error terms have been eliminated, we are left with an average that involves polynomial iterates. Next, we use an argument from [30] that allows us to pass to an extension of the system $(X,\mathcal{X},\mu,T)$. To be more precise, there exists an $\mathbb{R}$-action (see the definition in Section 2) $(Y,\mathcal{Y},v,S)$ and functions $\widetilde{f}_1,\widetilde{f}_2$, such that we have the equality

$$T^{\lfloor p_r(h)\rfloor}f_1\cdot T^{2\lfloor p_r(h)\rfloor}f_2=S_{p_r(h)}\widetilde{f}_1\cdot S_{2p_r(h)}\widetilde{f}_2.$$

This procedure can be done because the polynomial $p_r(h)$ has good equidistribution properties (which we analyze in the previous step) and thus the fractional parts of the finite sequence $\big(p_r(h)\big)_{0\leq h\leq L(r)}$ fall inside a small interval around 1 with the correct frequency. This is a necessary condition in order to use Proposition 3.1, which provides a bound for the inner average. To be more specific, we have the expression

$$\limsup_{r\to+\infty}\Big\|\mathop{\mathbb{E}}_{0\leq h\leq L(r)}\big(\Lambda_{w,b}(r+h)-1\big)S_{p_r(h)}\widetilde{f}_1\cdot S_{2p_r(h)}\widetilde{f}_2\Big\|_{L^2(\mu)}.$$

The inner average involves polynomials and can be bounded uniformly by the Gowers norm of the sequence $\Lambda_{w,b}(n)-1$ by Proposition 3.1 (modulo some constants and error terms that we ignore for the sake of this discussion). In particular, we have that the average in (19) is bounded by

$$\big\|\Lambda_{w,b}(n)-1\big\|_{U^s(r,r+L(r)]}$$

for some $s\in\mathbb{N}$. Finally, Theorem A implies that for sufficiently large values of r we have $\big\|\Lambda_{w,b}(n)-1\big\|_{U^s(r,r+L(r)]}$ is $o_w(1)$ uniformly in b. Finally, sending w to $+\infty$, we reach the desired conclusion.

This argument is quite simpler than the general case since it involves only one function. As we commented briefly, one extra complication is that we are dealing with a double averaging, unlike the model example. During the proof of Theorem 1.1, we will also need to split the functions a_{ij} into several distinct classes, which are handled with different methods. For example, the argument above works nicely for the function $t^{3/2}$ but has to be modified in the case of the function $\log^2 t$, because the latter cannot be approximated by polynomials of degree 1 or higher on our short intervals. Namely, the Taylor polynomial corresponding to $\log^2 t$ is constant and the previous method is orendered ineffective. Thus, we present an additional, more elaborate model example in Section 4, which exemplifies the possible cases that arise in the main proof.

1.4. **Open problems and further directions.** We expect that condition (11) in Theorem 1.6 can be relaxed significantly and still provide a multiple recurrence result. Motivated by [10, Theorem 2.3], we make the following conjecture.

Conjecture 1. *Let $a\in\mathcal{H}$ be a function of polynomial growth which satisfies*

$$\lim_{t\to+\infty}\big|a(t)-cp(t)\big|=+\infty \text{ for every } c\in\mathbb{R} \text{ and } p(t)\in\mathbb{Z}[t].$$

Then, for any $k\in\mathbb{N}$, measure-preserving system $(X,\mathcal{X},\mu,T)$ and set A of positive measure, the set

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

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

Comparing the assumptions on the function a to those in Theorem 1.6, we see that we are very close to establishing Conjecture 1. However, there are examples that our work does not encompass, such as the function $t^4+\log t$ or $t^2+\log\log(5t)$. In the setting of multiple recurrence along $\mathbb{N}$, the corresponding result was established in [10] and was generalized for more functions in [3]. In view of [3, Corollary B.3, Corollary B.4], we also make the following conjecture:

Conjecture 2. *Let $k\in\mathbb{N}$ and $a_1,\ldots,a_k\in\mathcal{H}$ be functions of polynomial growth. Assume that every non-trivial linear combination a of the functions $a_1,\ldots,a_k$, a, has the property*

$$\lim_{t\to+\infty}|a(t)-p(t)|=+\infty \text{ for all } p(t)\in\mathbb{Z}[t].$$

Then, for any measure-preserving system $(X,\mathcal{X},\mu,T)$ and set A of positive measure, the set

$$\{n\in\mathbb{N}:\ \mu\big(A\cap T^{-\lfloor a_1(n)\rfloor}A\cap\cdots\cap T^{-\lfloor a_k(n)\rfloor}A\big)>0\}$$

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

We remark that if one wants to also include functions that are essentially equal to a polynomial, then there are more results in this direction in [3], where it was shown that a multiple recurrence result for functions that are approximately equal to jointly-intersective polynomials is valid. Certainly, one would need to work with the sets $\mathbb{P}+1$ or $\mathbb{P}-1$ in this setting to transfer this result from $\mathbb{N}$ to the primes.

It is known that a convergence result along $\mathbb{N}$ with typical Cesàro averages cannot be obtained, if one works with the weaker conditions of the previous two conjectures. Indeed, the result would fail even for rotations on tori, because the corresponding equidistribution statement is false. The main approach employed in [3] was to consider a weaker averaging scheme than Cesàro averages. Using a different averaging method, one can impose some equidistribution assumption on functions that are not equidistributed in the standard sense. For instance, it is well-known that the sequence $(\log n)_{n\in\mathbb{N}}$ is not equidistributed mod 1 using Cesàro averages, but it is equidistributed under logarithmic averaging. Thus, it is natural to expect that an analog of Theorem 1.1 for other averaging schemes would allow someone to relax the conditions (7) and (8) in order to tackle the previous conjectures. A comparison result similar to Theorem 1.1 (but for other averaging schemes) appears to be a potential first step in this problem.

We expect that, under the same hypotheses, the analogous result in the setting of multiple commuting transformations will also hold. In particular, aside from the special cases established in [11], convergence results along $\mathbb{N}$ for Hardy sequences and commuting transformations are still open. For instance, it is unknown whether the averages in Theorem 1.5 converge when the functions a_i are linear combinations of fractional powers. In view of Theorem 1.1 and Theorem 1.2, any new result in this direction can be transferred to the setting of primes in a rather straightforward fashion, since conditions (7) and (8) are quite general to work with.

1.5. **Acknowledgements.** We thank Nikos Frantzikinakis for helpful discussions.

1.6. **Notational conventions.** Throughout this article, we denote with $\mathbb{N}=\{1,2,\ldots\}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ the sets of natural, integer, rational, real, and complex numbers respectively. We denote the one dimensional torus $\mathbb{T}=\mathbb{R}/\mathbb{Z}$, the exponential phases $e(t)=e^{2\pi it}$, 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 also let $\mathbf{1}_A$ denote the characteristic function of a set A and $|A|$ is its cardinality.

For any integer Q and $0 \leq a \leq Q-1$, 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 $h_1, \ldots, h_k$, we will use the notation $a_n = O_{h_1,\ldots,h_k}(b_n)$. 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 $a_N \ggg b_N$.[9] We use similar terminology and notation for functions on a real variable t.

Finally, for any sequence $(a(n))$, we employ the 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 will generally be non-integers. We shall avoid using this notation for the Cesàro averages.

2. Background

2.1. **Measure-preserving actions.** Let $(X, \mathcal{X}, \mu)$ be a Lebesgue probability space. A transformation $T : X \to X$ is *measure-preserving* if $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{X}$. It is called *ergodic* if all the T-invariant functions are constant. If T is invertible, then T induces a $\mathbb{Z}$-action on X by $(n, x) = T^n x$, for every $n \in \mathbb{Z}$ and $x \in X$.

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 $T_{g_1 g_2} = T_{g_1} \circ T_{g_2}$. For the purposes of this article, 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.

2.2. **Hardy fields.** Let $(\mathcal{B}, +, \cdot)$ denote the ring of germs at infinity of real-valued functions defined on a half-line $(t_0, +\infty)$. A sub-field $\mathcal{H}$ of $\mathcal{B}$ that is closed under differentiation is called a *Hardy field.* For any two functions $f, g \in \mathcal{H}$, with g not identically zero, the limit

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

exists in the extended line and thus we can always compare the growth rates of two functions in $\mathcal{H}$. In addition, every non-constant function in $\mathcal{H}$ is eventually monotone and has a constant sign eventually. We define below some notions that will be used repeatedly throughout the remainder of the paper.

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

Definition 2.1. *Let a be a function in $\mathcal{H}$. We say that the function a has* polynomial growth *if there exists a positive integer k such that $a(t) \ll t^k$. The smallest positive integer k for which this holds will be called the* degree *of a. The function a is called* sub-linear *if $a(t) \prec t$. It will be called* sub-fractional *if $a(t) \prec t^{\varepsilon}$, for all $\varepsilon > 0$. Finally, we will say that a is* strongly non-polynomial *if, for all positive integers k, we have that the functions $a(t)$ and t^k have distinct growth rates.*

Throughout the proofs in the following sections, we will assume that we have fixed a Hardy field $\mathcal{H}$. Some of the theorems impose certain additional assumptions on $\mathcal{H}$, but this is a byproduct of the arguments used to establish the case of convergence of Cesàro averages in [44] and we will not need to use these hypotheses in any of our arguments.

2.3. **Gowers uniformity norms on intervals of integers.** 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

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

and for $s \geq 2$,

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

A straightforward computation implies that

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

Here, the notation $\mathcal{C}$ denotes the conjugation map in $\mathbb{C}$, whereas for $\underline{\varepsilon} \in \{0,1\}^s$, $|\underline{\varepsilon}|$ is the sum of the entries of $\underline{\varepsilon}$ (the number of coordinates equal to 1).

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

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

for any function f on $\mathbb{Z}_N$ [27, Chapter 6]. For the purposes of this article, it will be convenient to consider similar expressions that are not necessarily defined only for functions in an abelian group $\mathbb{Z}_N$. Therefore, for any $s \geq 1$ and a finitely supported sequence $f(n), n \in \mathbb{Z}$, we define the *unnormalized Gowers uniformity norm*

$$\|f\|_{U^s(\mathbb{Z})} = \Big(\sum_{\underline{h}\in\mathbb{Z}^s}\sum_{n\in\mathbb{Z}}\prod_{\underline{\varepsilon}\in\{0,1\}^s}\mathcal{C}^{|\underline{\varepsilon}|}f(n+\underline{h}\cdot\underline{\varepsilon})\Big)^{\frac{1}{2^s}} \tag{20}$$

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

$$\|f\|_{U^s(I)} = \frac{\|f\cdot\mathbf{1}_I\|_{U^s(\mathbb{Z})}}{\|\mathbf{1}_I\|_{U^s(\mathbb{Z})}}. \tag{21}$$

First of all, observe that a simple change of variables in the summation in (21) 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 [27, Chapter 22, Proposition 11].

Lemma 2.2. *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 2.3. *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*

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

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 $\left\| u_X(n)\mathbf{1}_{a\ (Q)}(X+n) \right\|_{U^s[1,H]}$. 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

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

Applying the triangle inequality, we deduce that

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

However, it is immediate from (20) that the U^s-norm is invariant under multiplication by linear phases, for every $s \geq 2$. Therefore, we conclude that

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

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. In particular, Proposition 2.4 below implies that polynomial ergodic averages weighted by a sequence $(a(n))_{n\in\mathbb{N}}$ can be bounded in terms of the Gowers norm of a on the abelian group $\mathbb{Z}_{sN}$ for some positive integer s (that depends only on the degrees of the underlying polynomials).

Proposition 2.4. [17, Lemma 3.5] *Let $k, \ell \in \mathbb{N}$, $(X, \mathcal{X}, \mu, T_1, \dots, T_k)$ be a system of commuting $\mathbb{Z}$ actions, $p_{i,j} \in \mathbb{Z}[t]$ be polynomials for every $1 \leq i \leq k$, $1 \leq j \leq \ell$, $f_1, \dots, f_\ell \in L^\infty(\mu)$ 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*

(22)
$$\Big\| \mathop{\mathbb{E}}_{1\leq n\leq N} a(n) \cdot \prod_{j=1}^{\ell}\prod_{i=1}^{k} T_i^{p_{i,j}(n)} f_j \Big\|_{L^2(\mu)} \leq C_s \left(\left\| a \cdot \mathbf{1}_{[1,N]} \right\|_{U^s(\mathbb{Z}_{sN})} + \frac{\max\{1, \|a\|_{\ell^\infty[1,sN]}^{2s}\}}{N} \right).$$

Remark 4. *(i) The statement presented in* [17] *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 2.4 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 (22) 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 and the bound $\left\|\mathbf{1}_{[1,N]}\right\|_{U^s(\mathbb{Z}_{sN})} \leq 1$ implies that we can replace the norm in (22) 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 2.4 implies the following.

Corollary 2.5. *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*
(23)

$$\Big\| \underset{1\leq n\leq N}{\mathbb{E}} a(n) \cdot \prod_{j=1}^{\ell} \prod_{i=1}^{k} S_{i,p_{i,j}(n)} f_j \Big\|_{L^2(\mu)} \leq C_s \left(\left\| a \cdot \mathbf{1}_{[1,N]} \right\|_{U^s(\mathbb{Z}_{sN})} + \frac{\max\{1, \|a\|_{\ell^\infty[1,sN]}^{2s}\}}{N} \right).$$

2.4. **Number theoretic tools.** The following lemma is a standard consequence of the prime number theorem and the sparseness of prime powers (actually, we use this argument in the proof of Corollary 2.8 below). For a proof, see, for instance, [27, Chapter 25].

Lemma 2.6. *For any bounded sequence $(a(n))_{n\in\mathbb{N}}$ in a normed space, we have*

$$\lim_{N\to+\infty} \Big\| \frac{1}{\pi(N)} \sum_{p\in\mathbb{P}:\, p\leq N} a(p) - \frac{1}{N} \sum_{n=1}^{N} \Lambda(n) a(n) \Big\| = 0. \tag{24}$$

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

For the modified von Mangoldt function, we have the following deep theorem, which was recently established in [36].

Theorem A. [36, 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 in terms of w, we have that*

$$\|\Lambda_{w,b} - 1\|_{U^s(N, N+L(N)]} = o_w(1) \tag{25}$$

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

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 A in conjunction with an application of the van der Corput inequality d times until the polynomial phase is eliminated. Alternatively, one can use Proposition 2.4 for a rotation on the torus $\mathbb{T}$ to carry out the reduction to Theorem A.[10] We omit its proof.

[10]Evidently, both statements rely on similar complexity reduction arguments, though Proposition 2.4 is stated in much larger generality involving numerous polynomials.

Lemma 2.7. *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*

$$\max_{\substack{1\leq b\leq W\\ (b,W)=1}} \sup_{\substack{p\in\mathbb{R}[t]\\ \deg p=d}} \Big| \underset{N\leq n\leq N+L(N)}{\mathbb{E}} \big(\Lambda_{w,b}(n)-1\big)e(p(n))\Big| = o_w(1). \tag{26}$$

for every N large enough in terms of w.

Remark 5. (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* [36]*, 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 [28], since $L(N)$ will be chosen to grow sufficiently fast in our applications.

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

$$\pi(x+y,q,a)-\pi(x,q,a) \leq \frac{2y}{\phi(q)\log(\frac{y}{q})} \tag{27}$$

for every $x \geq y > q$.

While we referred to this as the Brun-Titchmarsh inequality, the previous theorem was established in [37] 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. This follows easily from the previous theorem and a standard partial summation argument.

Corollary 2.8. *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

$$\Big|\theta(x,q,a)-\psi(x,q,a)\Big| \leq \sum_{p^k\leq x\colon p\in\mathbb{P},k\geq 2} \log p \leq x^{1/2}\log x, \tag{28}$$

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

$$\theta(x+y,q,a)-\theta(x,q,a)\leq \frac{2y\log x}{\phi(q)\log(\frac{y}{q})}+O(\frac{y}{\log x})+O(1). \tag{29}$$

Consequently, if we combine (28) and (29), 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 6. *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.*

2.5. **Quantitative equidistribution mod 1.**

Definition 2.9. *Let $(x_n)_{n\in\mathbb{N}}$ be a real valued sequence. We say that $(x_n)_{n\in\mathbb{N}}$ is*

- equidistributed *mod* 1 *if for all $0\leq a<b\leq 1$, we have*

$$\lim_{N\to+\infty}\frac{\big|\{n\in\{1,\dots,N\}:\ \{x_n\}\in[a,b)\}\big|}{N}=b-a. \tag{30}$$

- well distributed *mod* 1 *if for all $0\leq a<b\leq 1$, we have*

$$\lim_{N\to+\infty}\frac{\big|\{n\in\{1,\dots,N\}:\ \{x_{k+n}\}\in[a,b)\}\big|}{N}=b-a,\ \text{uniformly in } k=0,1,\dots. \tag{31}$$

In the case of polynomial sequences, 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. On the other hand, for polynomials with at least one non-constant irrational coefficient, we have the following theorem.

Theorem C (Weyl)**.** *Let $p \in \mathbb{R}[t]$ be a polynomial with at least one non-constant irrational coefficient. Then, the sequence $(p(n))_{n\in\mathbb{N}}$ is well-distributed mod 1.*

This theorem is classical and for a proof, we refer the reader to [32, Chapter 1, Theorem 3.2].[11] In the case of Hardy field functions, we have a complete characterization of equidistribution modulo 1 due to Boshernitzan. We recall here [4, Theorem 1.3].

Theorem D (Boshernitzan)**.** *Let $a \in \mathcal{H}$ be a function of polynomial growth. Then, the sequence $(a(n))_{n\in\mathbb{N}}$ is equidistributed mod 1 if and only if $|a(t) - p(t)| \succ \log t$ for every $p \in \mathbb{Q}[t]$.*

This theorem explains the assumptions in Theorem 1.1 and, in particular, condition (7). Indeed, since we need equidistribution assumptions for our method to work, this condition appears to be vital. We will invoke Boshernitzan's theorem only in the case of sub-fractional functions. Indeed, we will investigate the equidistribution properties of fast-growing functions by studying their exponential sums in short intervals. This leads to a proof of the previous theorem indirectly, at least in the case that the function involved is not sub-fractional.

For our purposes, we will need a quantitative version of the equidistribution phenomenon. For a finite sequence of real numbers $(u_n)_{1\leq n\leq N}$ and an interval $[a, b] \subseteq [0, 1]$, we define the *discrepancy* of the sequence u_n with respect to $[a, b]$ by

$$\Delta_{[a,b]}(u_1, \dots, u_N) = \left| \frac{\big|\{n \in \{1, \dots, N\} \colon \{u_n\} \in [a, b]\}\big|}{N} - (b - a) \right|. \tag{32}$$

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 $0 \leq a \leq b \leq 1$. 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 [32]. 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 [32, Chapter 2, Theorem 2.5].[12]

Theorem E (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*

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

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

[11] While this theorem concerns the case of equidistribution, the more general result follows easily by a straightforward adaptation of van der Corput's difference theorem to the case of well-distribution. The authors of [32] discuss this in the notes of Section 5 in Chapter 1.

[12] 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.

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

$$\Delta_{[a,b]}(u_1,\ldots,u_N) \leq C\Big(\frac{1}{M} + \sum_{m=1}^{M}\frac{1}{m}\Big|\frac{1}{N}\sum_{n=1}^{N} e(mu_n)\Big|\Big) \tag{33}$$

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

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 are "major-arc". 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, which will not satisfy such a major-arc condition. To this end, we require the following lemma.

Lemma 2.10. *Let* $0 < \delta < 1$ *and* $d \in \mathbb{N}$. *There exists a positive constant* C *depending only on* d, *such that if* $p(x) = a_d x^d + \cdots + a_1 x + a_0$ *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* $N^k \left\|qa_k\right\|_{\mathbb{T}} \leq \delta^{-C}$.

Note 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). For a proof of this lemma, see [23, Proposition 4.3], where a more general theorem is established in the setting of nilmanifolds as well.

2.6. **Nilmanifolds and correlation sequences.** Let G be a nilpotent Lie group with nilpotency degree s and let Γ be a discrete and cocompact subgroup. The space $X = G/\Gamma$ is called an s-step *nilmanifold*. The group G acts on the space X by left multiplication and the measure on X that is invariant under this action is called the *Haar measure* of X, which we shall denote by m_X.

Given a sequence of points $x_n \in X$, we will say that the sequence x_n is *equidistributed* on X, if for any continuous function $F : X \to \mathbb{C}$ we have that

$$\lim_{N\to+\infty}\frac{1}{N}\sum_{n=1}^{N} F(x_n) = \int F\, dm_X.$$

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.

Let g be any element on the group G. Then, for any $x \in X$, the closed orbit of the action of g on x will be denoted by $\overline{(g^{\mathbb{Z}}x)}$. It is known that this set is a subnilmanifold of $X = G/\Gamma$ and that the sequence $g^n x$ is equidistributed in the subnilamnifold $\overline{(g^{\mathbb{Z}}x)}$ (see, for example, [27, Chapter 11, Theorem 9]).

We now present the following definition for nilsequences in several variables.

Definition 2.11. *Let* k, s *be positive integers and let* $X = G/\Gamma$ *be an* 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*

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

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

The main tool that we will need is an approximation of general nilsequences by multicorrelation sequences in the ℓ^∞-sense. The following lemma is established in [15, Proposition 4.2].

Lemma 2.12. *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*

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

with $\ell_j = s!/j$ satisfies

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

Comment. The definition of nilsequences used in [15] 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.

3. 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 [29] for Cesàro averages, so we adapt its proof to the setting of short intervals.

Proposition 3.1. *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*

$$\lim_{\delta \to 0^+} \lim_{N \to +\infty} \frac{|\{N \leq n \leq N + L(N) : \ \{p_{i,j,N}(n)\} \in [1-\delta, 1)\}|}{L(N)} = 0, \tag{34}$$

we have that for any $0 < \delta < 1$ and functions $f_1, \ldots, f_\ell \in L^\infty(\mu)$

$$\Big\| \mathop{\mathbb{E}}_{N \leq n \leq N+L(N)} \big(\Lambda_{w,b}(n) - 1\big) \prod_{j=1}^{\ell} \prod_{i=1}^{k} T_i^{\lfloor p_{i,j,N}(n) \rfloor} f_j \Big\|_{L^2(\mu)} \ll_{k,\ell,d}$$
$$\frac{1}{\delta^{k\ell}} \Big(\|\Lambda_{w,b}(n) - 1\|_{U^s(N, N+sL(N)]} + o_w(1) \Big) + o_\delta(1)(1 + o_w(1)),$$

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 [29]) 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 $f_1, \dots, f_\ell$ are bounded functions on X, we define the Y-extensions of f_j, setting for every element $(a_{1,1}, \dots, a_{k,1}, a_{1,2}, \dots, a_{k,2}, \dots, a_{1,\ell}, \dots, a_{k,\ell}) \in [0,1)^{k\ell}$:

$$\hat{f}_j(x, a_{1,1}, \dots, a_{k,1}, a_{1,2}, \dots, a_{k,2}, \dots, a_{1,\ell}, \dots, a_{k,\ell}) = f_j(x), \ \ 1 \leq j \leq \ell;$$

and we also define the function

$$\hat{f}_0(x, a_{1,1}, \dots, a_{k,1}, a_{1,2}, \dots, a_{k,\ell}) = 1_{[0,\delta]^{k\ell}}(a_{1,1}, \dots, a_{k,1}, a_{1,2}, \dots, a_{k,\ell}).$$

For every $N \leq n \leq N + L(N)$, we consider the functions (on the original space X)

$$b_N(n) := (\prod_{i=1}^{k} T_i^{[p_{i,1,N}(n)]})f_1 \cdot \dots \cdot (\prod_{i=1}^{k} T_i^{[p_{i,\ell,N}(n)]})f_\ell$$

as well as the functions

$$\tilde{b}_N(n) := \hat{f}_0 \cdot (\prod_{j=1}^{\ell}\prod_{i=1}^{k} S_{i,\delta_{j1}\cdot p_{i,1,N}(n)})\hat{f}_1 \cdot \dots \cdot (\prod_{j=1}^{\ell}\prod_{i=1}^{k} S_{i,\delta_{j\ell}\cdot p_{i,\ell,N}(n)})\hat{f}_\ell$$

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

$$b'_N(n)(x) := \int_{[0,1)^{k\ell}} \tilde{b}_N(n)(x, a_{1,1}, \dots, a_{k,1}, a_{1,2}, \dots, a_{k,2}, \dots, a_{1,\ell}, \dots, a_{k,\ell})\, d\lambda^{k\ell},$$

where the integration is with respect to the variables $a_{i,j}$.

Using the triangle and Cauchy-Schwarz inequalities, we have

$$(35) \quad \delta^{k\ell}\Big\| \mathop{\mathbb{E}}_{N\leq n\leq N+L(N)} \big(\Lambda_{w,b}(n) - 1\big) b_N(n)\Big\|_{L^2(\mu)} \leq$$
$$\Big\| \mathop{\mathbb{E}}_{N\leq n\leq N+L(N)} \big(\Lambda_{w,b}(n){-}1\big)\cdot(\delta^{k\ell} b_N(n){-}b'_N(n))\Big\|_{L^2(\mu)} {+} \Big\| \mathop{\mathbb{E}}_{N\leq n\leq N+L(N)} \big(\Lambda_{w,b}(n){-}1\big)\tilde{b}_N(n)\Big\|_{L^2(\nu)}.$$

Using Proposition 2.4, 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

(36)
$$\Big\| \mathop{\mathbb{E}}_{N\leq n\leq N+L(N)} \big(\Lambda_{w,b}(n) - 1\big)\tilde{b}_N(n)\Big\|_{L^2(\nu)} \leq C_s \left(\|\Lambda_{w,b} - 1\|_{U^s(N,N+sL(N)]} + o_N(1)\right),$$

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

$$\Big\| \mathop{\mathbb{E}}_{N\leq n\leq N+L(N)} \big(\Lambda_{w,b}(n) - 1\big) \cdot (\delta^{k\ell} b_N(n) - b'_N(n))\Big\|_{L^2(\mu)}$$

in (35). For every $x \in X$ and $N \leq n \leq N + L(N)$, we have

$$\Big|\delta^{k\ell} b_N(n)(x) - b'_N(n)(x)\Big| =$$
$$\left| \int_{[0,\delta]^{k\ell}} \left(\prod_{j=1}^{\ell} f_j(\prod_{i=1}^{k} T_i^{[p_{i,j,N}(n)]}x) - \prod_{j=1}^{\ell} f_j(\prod_{i=1}^{k} T_i^{[p_{i,j,N}(n)+a_{i,j}]}x) \right) d\lambda^{k\ell} \right|.$$

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 $1 \leq i \leq k$, $1 \leq j \leq \ell$. To deal with the possible case where $\{p_{i,j,N}(n)\} \geq 1 - \delta$ for at least one of our polynomials, we define, for every $1 \leq i \leq k$, $1 \leq j \leq \ell$, 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

$$\mathbf{1}_{E^{1,1}_{\delta,N}\cup\ldots\cup E^{1,\ell}_{\delta,N}\cup E^{2,1}_{\delta,N}\cup\ldots\cup E^{k,\ell}_{\delta,N}} \leq \sum_{(i,j)\in[1,k]\times[1,\ell]} \mathbf{1}_{E^{i,j}_{\delta,N}}$$

and that $\mathbf{1}_{E^{i,j}_{\delta,N}}(n) = \mathbf{1}_{[1-\delta,1)}(\{p_{i,j,N}(n)\})$, we infer that

$$\left|\delta^{k\ell} b_N(n)(x) - b'_N(n)(x)\right| \leq 2\delta^{k\ell} \sum_{(i,j)\in[1,k]\times[1,\ell]} \mathbf{1}_{[1-\delta,1)}(\{p_{i,j,N}(n)\})$$

for every $x \in X$. In view of the above, using the inequality $|\Lambda_{w,b}(n) - 1| \leq \Lambda_{w,b}(n) + 1$, we deduce that

$$\underset{N\leq n\leq N+L(N)}{\mathbb{E}} \left|\big(\Lambda_{w,b}(n) - 1\big)\right| \cdot \mathbf{1}_{[1-\delta,1)}(\{p_{i,j,N}(n)\}) \leq$$
$$\underset{N\leq n\leq N+L(N)}{\mathbb{E}} \big(\Lambda_{w,b}(n) - 1\big) \cdot \mathbf{1}_{[1-\delta,1)}(\{p_{i,j,N}(n)\}) + 2 \underset{N\leq n\leq N+L(N)}{\mathbb{E}} \mathbf{1}_{[1-\delta,1)}(\{p_{i,j,N}(n)\}) \leq$$
$$\underset{N\leq n\leq N+L(N)}{\mathbb{E}} \big(\Lambda_{w,b}(n) - 1\big) \cdot \mathbf{1}_{[1-\delta,1)}(\{p_{i,j,N}(n)\}) + 2 \cdot \frac{|E^{i,j}_{\delta,N}|}{L(N)}.$$

Since each polynomial $p_{i,j,N}$ satisfies (34) 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

$$\underset{N\leq n\leq N+L(N)}{\mathbb{E}} \big(\Lambda_{w,b}(n) - 1\big) \cdot \mathbf{1}_{[1-\delta,1)}(\{p_{i,j,N}(n)\})$$

goes to zero as $N \to \infty$, then $w \to \infty$ and finally $\delta \to 0^+$. To this end, it suffices to show

$$\underset{N\leq n\leq N+L(N)}{\mathbb{E}} \big(\Lambda_{w,b}(n) - 1\big)e^{2\pi i m p_{i,j,N}(n)} \to 0$$

as $N \to \infty$ and then $w \to \infty$ for all $m \in \mathbb{Z} \setminus \{0\}$,[13] which follows from Lemma 2.7. □

4. 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 4.6 below. We will also present an example that illustrates the main points in the proof of Theorem 1.1 in the following section.

4.1. Details on the proof.

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 demonstrate 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

[13] 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 $q_1(t) + c - \varepsilon \leq f(t) \leq q_2(t) + c + \varepsilon$. We use this for the function $f = \mathbf{1}_{[1-\delta,1)}$.

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 2.8.

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 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 $\Lambda_{w,b}(n)$.

4.2. **A model example.** We sketch the main steps in the case of the ergodic averages

$$\frac{1}{N} \sum_{n=1}^{N} \big(\Lambda_{w,b}(n)-1\big) T^{\lfloor n \log n \rfloor} f_1 \cdot T^{\lfloor an^2+\log n \rfloor} f_2 \cdot T^{\lfloor \log^2 n \rfloor} f_3. \tag{37}$$

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 1.1), 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

$$\mathop{\mathbb{E}}_{1 \leq r \leq R} \Big\| \mathop{\mathbb{E}}_{r \leq n \leq r+L(r)} \big(\Lambda_{w,b}(n)-1\big) T^{\lfloor n \log n \rfloor} f_1 \cdot T^{\lfloor an^2+\log n \rfloor} f_2 \cdot T^{\lfloor \log^2 n \rfloor} f_3 \Big\|_{L^2(\mu)} = 0. \tag{38}$$

This reduction is the content of Lemma 5.1. 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

$$\Big|\frac{(n-r)^3}{6\xi_{1,n,r}^2}\Big| \leq \frac{r^{3\cdot 0.65}}{6r^2} \ll 1,$$

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

$$\lfloor \log^2 n\rfloor = \lfloor \log^2 r\rfloor,$$

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 2.8, though we will ignore that in this exposition. In conclusion, we can rewrite the expression in (38) as

$$\underset{1\leq r\leq R}{\mathbb{E}}\Big\|\underset{r\leq n\leq r+L(r)}{\mathbb{E}}\big(\Lambda_{w,b}(n)-1\big)T^{\lfloor n\log n\rfloor}f_1\cdot T^{\lfloor an^2+\log n\rfloor}f_2\cdot T^{\lfloor \log^2 r\rfloor}f_3\Big\|_{L^2(\mu)}+O(\varepsilon). \tag{39}$$

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 4.4 using Lemma 2.10 and Theorem E. As a baby case, we show the following estimate for some simple trigonometric averages:

$$\Big|\underset{r\leq n\leq r+L(r)}{\mathbb{E}} e\Big(\frac{(n-r)^2}{2r}\Big)\Big| \leq \frac{1}{\log^A r}$$

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

$$\Big\|\frac{q_r}{2r}\Big\|_{\mathbb{T}} \leq \frac{\log^{O(A)} r}{(L(r))^2}.$$

Certainly, 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

$$\lfloor n \log n \rfloor = \left\lfloor r \log r + (\log r + 1)(n - r) + \frac{(n-r)^2}{2r} \right\rfloor$$

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 (39) as

(40)

$$\mathop{\mathbb{E}}_{1\leq r\leq R} \Big\| \mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} \big(\Lambda_{w,b}(n) - 1\big) T^{\lfloor p_r(n) \rfloor} f_1 \cdot T^{\lfloor an^2 + \log n \rfloor} f_2 \cdot T^{\lfloor \log^2 r \rfloor} f_3 \Big\|_{L^2(\mu)} + O(\varepsilon) + o_R(1),$$

where $p_r(n) = r \log r + (\log r + 1)(n - r) + \frac{(n-r)^2}{2r}$.

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

$$\max_{1\leq m\leq M} \Big| \mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} e\big(m(an^2 + \log r)\big) \Big| = o_r(1). \tag{41}$$

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 E implies that

$$\frac{|\mathcal{B}_{r,\varepsilon}|}{L(r)} \ll 2\varepsilon + \frac{1}{M} + \sum_{m=1}^{M} \frac{1}{m} \Big| \mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} e\big(m(ar^2 + \log r)\big) \Big|, \tag{42}$$

so that $|\mathcal{B}_{r,\varepsilon}| \ll (\varepsilon + o_r(1))L(r)$. Additionally, we will need to estimate

$$\frac{1}{L(r)} \sum_{r\leq n\leq r+L(r)} \Lambda_{w,b}(n) \mathbf{1}_{\mathcal{B}_r}(n),$$

which will arise when we apply the triangle inequality to bound the contribution of the set $\mathcal{B}_r$. However, we have that

$$\max_{1\leq m\leq M} \Big| \mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} \Lambda_{w,b}(n) e\big(m(an^2 + \log r)\big) \Big| = o_w(1) + o_r(1), \tag{43}$$

which can be seen by splitting $\Lambda_{w,b}(n) = (\Lambda_{w,b}(n) - 1) + 1$, applying the triangle inequality and using Lemma 2.7 and (41), respectively, to treat the resulting exponential averages. In view of this, we can apply the Erdős-Turán inequality (Theorem E) for the probability measure

$$\nu(S) = \frac{\sum_{r\leq n\leq r+L(r)} \Lambda_{w,b}(n) \delta_{\{an^2 + \log r\}}(S)}{\sum_{r\leq n\leq r+L(r)} \Lambda_{w,b}(n)}$$

as well as Corollary 2.8 (to bound the sum in the denominator) to conclude that

$$\frac{1}{L(r)} \sum_{r\leq n\leq r+L(r)} \Lambda_{w,b}(n)\mathbf{1}_{\mathcal{B}_r}(n) \ll \varepsilon + o_w(1)\log\frac{1}{\varepsilon} + o_r(1),$$

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 that turn out to be satisfactory, we can rewrite the expression in (40) as

$$\underset{1\leq r\leq R}{\mathbb{E}} \Big\| \underset{r\leq n\leq r+L(r)}{\mathbb{E}} \big(\Lambda_{w,b}(n)-1\big) T^{\lfloor p_r(n)\rfloor} f_1 \cdot T^{\lfloor an^2+\log r\rfloor} f_2 \cdot T^{\lfloor \log^2 r\rfloor} f_3 \Big\|_{L^2(\mu)}. \tag{44}$$

Now, the iterates satisfy the assumptions of Proposition 3.1. 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 A to show that the average

$$\underset{1\leq r\leq R}{\mathbb{E}} \|\Lambda_{w,b} - 1\|_{U^s(r,r+L(r)]}$$

converges to 0, which leads us to our desired conclusion.

4.3. **Some preparatory lemmas.** Let us fix a Hardy field $\mathcal{H}$. 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 [9, Lemma 2.1], as well as [44, Proposition A.1].

Lemma 4.1. *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,*

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

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 $a'(t)$.

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 4.2. [44, Proposition A.2] *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 7. *The proof of Proposition 4.2 in* [44] *establishes the fact that if a satisfies the previous hypotheses, then the derivatives of a always satisfy the conditions of Lemma 4.1.*

This proposition is the main tool used 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

$$|a^{(k)}(t)|^{-\frac{1}{k}} \prec L(t) \prec |a^{(k+1)}(t)|^{-\frac{1}{k+1}} \tag{45}$$

for some sufficiently large $k \in \mathbb{N}$ (large enough so that the inequalities in Proposition 4.2 hold). In particular, this implies that $\lim\limits_{t\to+\infty} a^{(k+1)}(t) \to 0$.

Then, we can use the Taylor expansion around the point N to 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 $0 \le h \le L(N)$. However, we observe that

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

where we used the fact that $|a^{(k+1)}(t)| \to 0$ monotonically (since $a^{(k+1)}(t) \in \mathcal{H}$). 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.

Proposition 4.3. *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*

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

where we adopt the convention that $|a^{(k)}(t)|^{-\frac{1}{k}}$ 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 4.1 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

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

for k large enough, due to Proposition 4.2. 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 $L(t) \prec |a^{(d+1)}(t)|^{-\frac{1}{d+1}}$, then the claim holds for $k = d$, since $|a^{(d)}(t)|^{-\frac{1}{d}} \prec 1 \prec L(t)$.

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

$$t^{1-\varepsilon} \prec |a^{(k+1)}(t)|^{-\frac{1}{k+1}} \tag{46}$$

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 4.1 (see also Remark 7). 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 8. *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 $0 \le h \le \frac{N}{\log N}$ 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.

4.4. **Eliminating the error terms in the approximations.** In the previous subsection, 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

$$|a^{(k)}(t)|^{-\frac{1}{k}} \prec L(t) \prec |a^{(k+1)}(t)|^{-\frac{1}{k+1}} \tag{47}$$

then, for any $0 \le h \le L(N)$, 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 3.1, we will need to eliminate the error terms $\varepsilon_{N,h}$. We will consider three distinct cases, which are tackled using somewhat different arguments.

4.4.1. *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 (47), which, as we saw previously, are sufficient to have a Taylor approximation in the interval $[N, N+L(N)]$.

Proposition 4.4. *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*

$$1 \lll |a^{(k)}(t)|^{-\frac{1}{k}} \lll L(t) \lll |a^{(k+1)}(t)|^{-\frac{1}{k+1}} \tag{48}$$

and such that the function $a^{(k+1)}(t)$ *converges to 0 monotonically. Then, for* N *large enough, we have that, for all* $0 \leq c \leq d < 1$,

$$\frac{\big|\{n \in [N, N+L(N)] \colon a(n) \in [c,d]\}\big|}{L(N)} = |d-c| + O_A(L(N)\log^{-A} N).^{14} \tag{49}$$

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

$$L(t)\big|a^{(k)}(t)\big|^{\frac{1}{k}} \gg t^{\varepsilon_1} \text{ and } L(t)\big|a^{(k+1)}(t)\big|^{\frac{1}{k+1}} \ll t^{-\varepsilon_2}. \tag{50}$$

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

$$a(N+h) = a(N) + ha'(N) + \cdots + \frac{h^k a^{(k)}(N)}{k!} + \frac{h^{k+1}a^{(k+1)}(\xi_h)}{(k+1)!}, \quad \text{for some } \xi_h \in [N, N+h], \tag{51}$$

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,

$$\max_{0 \leq h \leq L(N)} |\theta_N(h)| \leq \left|\frac{a^{(k+1)}(N)}{(k+1)!}\right| (L(N))^{k+1} = \theta_N, \tag{52}$$

and the quantity θ_N is strongly dominated by the constant 1 due to (50). 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

$$p_N(h), \text{ where } 0 \leq h \leq L(N).$$

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 E for the finite sequence $(p_N(h))_{0 \leq h \leq L(N)}$ to deduce that

$$\Delta_{[c,d]}\Big(\big(p_N(h)\big)_{0 \leq h \leq L(N)}\Big) \leq \frac{C}{\lfloor \log^A N \rfloor} + C \sum_{m=1}^{\lfloor \log^A N \rfloor} \frac{1}{m} \Big| \underset{0 \leq h \leq L(N)}{\mathbb{E}} e(mp_N(h)) \Big|, \tag{53}$$

where C is an absolute constant. We claim that for every $1 \leq m \leq \lfloor \log^A N \rfloor$, we have that

$$\Big| \underset{0 \leq h \leq L(N)}{\mathbb{E}} e(mp_N(h)) \Big| \leq \frac{1}{\log^A N}, \tag{54}$$

[14] 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 (48), though this will not be any more useful for our purposes.

provided that N is sufficiently large. Indeed, assume for the sake of contradiction that there exists $1 \leq m_0 \leq \lfloor \log^A N \rfloor$, such that

$$\Big| \mathop{\mathbb{E}}_{0 \leq h \leq L(N)} e(m_0 p_N(h)) \Big| > \frac{1}{\log^A N}. \tag{55}$$

The leading coefficient of $m_0 p_N(h)$ is equal to

$$\frac{m_0 a^{(k)}(N)}{k!}.$$

Then, Lemma 2.10 implies that there exists a constant C_k (depending only on k) an integer q satisfying $|q| \leq \log^{C_k A} N$ and such that

$$\Big\| q \cdot \frac{m_0 a^{(k)}(N)}{k!} \Big\|_{\mathbb{T}} \leq \frac{\log^{C_k A} N}{\lfloor L(N) \rfloor^k}.$$

The number qm_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 (55) with the absolute value, which readily implies that

$$\Big| q \cdot \frac{m_0 a^{(k)}(N)}{k!} \Big| \leq \frac{\log^{C_k A} N}{\lfloor L(N) \rfloor^k} \implies \lfloor L(N) \rfloor^k \big| a^{(k)}(N) \big| \leq k! \log^{C_k A} N.$$

However, this implies that $L(t)$ cannot strongly dominate the function $\big(a^{(k)}(t)\big)^{-\frac{1}{k}}$, which is a contradiction due to our hypothesis.

We have established that for every $1 \leq m \leq \lfloor \log^A N \rfloor$ and large N, inequality (54) holds. Substituting this in (53), we deduce that

$$\Delta_{[c,d]}\Big(\big(p_N(h)\big)_{0 \leq h \leq L(N)} \Big) \leq \frac{C}{\lfloor \log^A N \rfloor} + C \sum_{m=1}^{\lfloor \log^A N \rfloor} \frac{1}{m \log^A N},$$

which implies that

$$\Delta_{[c,d]}\Big(\big(p_N(h)\big)_{0 \leq h \leq L(N)} \Big) \ll \frac{A \log \log N}{\log^A N}.$$

In particular, since A was arbitrary, we get

$$\Delta_{[c,d]}\Big(\big(p_N(h)\big)_{0 \leq h \leq L(N)} \Big) \ll_A \frac{1}{\log^A N}. \tag{56}$$

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 (52) 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 (51) 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.

4.4.2. *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 (see Definition 2.1) 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 4.5. *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

$$a(N+h) = a(N) + ha'(\xi_h) \tag{57}$$

for some $\xi_h \in [N, N+h]$. In addition, since $a'(t)$ converges to 0 monotonically, we have

$$|ha'(\xi_h)| \le L(N)a'(N) \ll N^{1-\delta}a'(N) \lll 1,$$

where the last inequality follows from Lemma 4.1 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 $h \in [0, L(N)]$.[15]

The sequence $a(n)$ is equidistributed mod 1 by Theorem D, 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 $R \ge R_0$, 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 (57) 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

[15]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.

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 D), this is practically true for almost all N, if we take ε sufficiently small.

4.4.3. *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 4.4 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.

Proposition 4.6. *Let k,d be positive integers, let $0<\varepsilon<1/2$ be a real number and let $w\in\mathbb{N}$. We define $W=\prod_{p\in\mathbb{P}:\,p\le w}p$ and let $1\le b\le W$ 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*

$$t^{\frac{5}{8}}\lll L(t)\lll t.^{16}$$

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

$$\frac{1}{L(r)}\sum_{r\le n\le r+L(r)}\Lambda_{w,b}(n)\mathbf{1}_{\mathcal{B}_{r,\varepsilon}}(n)\ll_{k,d}\varepsilon+o_w(1)\log\frac{1}{\varepsilon}+o_r(1). \tag{58}$$

Remark 9. *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 1.1. On the other hand, the $o_w(1)$ term is the same as the one in 2.7 and depends on the degree d of the polynomials, which will be fixed in applications.*

Proof of Proposition 4.6. Fix an index $1\le i\le k$ 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 4.1, we deduce that

$$\max_{0\le h\le L(r)}|x_i(r+h)-x_i(r)|\le L(r)|x_i'(r)|\lll 1.$$

[16] See the notational conventions for the definition of $\lll$.

In particular, there exists $\delta_0 > 0$ depending only on the functions $a_1, \dots, a_k$ and $L(t)$, such that

$$\max_{0 \le h \le L(r)} |x_i(r+h) - x_i(r)| \ll r^{-\delta_0} \tag{59}$$

for all $1 \le i \le k$. 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

$$\mathcal{B}_{i,r,\varepsilon} = \{n \in [r, r+L(r)] \colon \{p_i(n) + x_i(r)\} \in [0,\varepsilon] \cup [1-\varepsilon, 1)\} \tag{60}$$

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 (58). 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, [32, Theorem 5.2, Chapter 1]) to conclude that

$$\max_{1 \le m \le M} \Big| \mathop{\mathbb{E}}_{r \le n \le r+L(r)} e\big(m(p_i(n) + x_i(r))\big) \Big| = o_r(1),$$

for all r sufficiently large in terms of M, as we have assumed to be the case.[17] On the other hand, Lemma 2.7 implies that

$$\max_{1 \le m \le M} \Big| \mathop{\mathbb{E}}_{r \le n \le r+L(r)} \big(\Lambda_{w,b}(n) - 1\big) e\big(m(p_i(n) + x_i(r))\big) \Big| = o_w(1)$$

for r sufficiently large in terms of w. Combining the last two bounds, we deduce that

$$\max_{1 \le m \le M} \Big| \mathop{\mathbb{E}}_{r \le n \le r+L(r)} \Lambda_{w,b}(n) e\big(m(p_i(n) + x_i(r))\big) \Big| = o_w(1) + o_r(1). \tag{61}$$

Since we have estimates on the exponential sums weighted by $\Lambda_{w,b}(n)$, we can now make the passage to (58). To this end, we apply Theorem E for the probability measure

$$\nu(S) = \frac{\sum\limits_{r \le n \le r+L(r)} \Lambda_{w,b}(n) \delta_{\{p_i(n) + x_i(r)\}}(S)}{\sum\limits_{r \le n \le r+L(r)} \Lambda_{w,b}(n)}.$$[18]

Setting

$$S_r = \sum_{r \le n \le r+L(r)} \Lambda_{w,b}(n)$$

for brevity, we conclude that

$$\frac{\sum\limits_{r \le n \le r+L(r)} \Lambda_{w,b}(n) \delta_{\{p_i(n)+x_i(r)\}}\big([0,\varepsilon] \cup [1-\varepsilon, 1)\big)}{S_r} \ll 2\varepsilon + \frac{1}{M} + \sum_{m=1}^{M} \frac{1}{m} \Big| \frac{1}{S_r} \sum_{r \le n \le r+L(r)} \Lambda_{w,b}(n) e\big(m(p_i(n) + x_i(r))\big) \Big|, \tag{62}$$

[17] 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.
[18] The denominator is non-zero if r is large enough.

where the implied constant is absolute. Applying the bounds in (61) and recalling the definition of $\mathcal{B}_{i,r,\varepsilon}$, we conclude that

$$\sum_{r\leq n\leq r+L(r)} \Lambda_{w,b}(n)\mathbf{1}_{\mathcal{B}_{i,r,\varepsilon}}(n) \ll \Big(\varepsilon+\frac{1}{M}\Big)S_r + \sum_{m=1}^{M}\frac{L(r)}{m}(o_w(1)+o_r(1)) \tag{63}$$
$$\ll \varepsilon S_r + L(r)\big(o_w(1)+o_r(1)\big)\log\frac{1}{\varepsilon},$$

since $M=\lfloor \varepsilon^{-1}\rfloor$. Finally, we bound S_r by applying Corollary 2.8 to conclude that

$$S_r = \frac{\phi(W)}{W}\sum_{\substack{Wr+b\leq n\leq Wr+b+WL(r)\\ n\equiv b\ (W)}} \Lambda(n) \leq \frac{\phi(W)}{W}\Big(\frac{2WL(r)\log r}{\phi(W)\log\big(\frac{L(r)}{W}\big)}+ \tag{64}$$
$$O\Big(\frac{L(r)}{\log r}\Big)+O(r^{1/2}\log r)\Big) \ll L(r)(1+o_r(1)),$$

where we used the fact that $L(r)\gg t^{5/8}$ to bound the first fraction by an absolute constant. Applying this in (63), we conclude that

$$\frac{1}{L(r)}\sum_{r\leq n\leq r+L(r)} \Lambda_{w,b}(n)\mathbf{1}_{\mathcal{B}_{i,r,\varepsilon}}(n) \ll \varepsilon(1+o_r(1)) + \big(o_w(1)+o_r(1)\big)\log\frac{1}{\varepsilon}.$$

Finally, we recall that $\mathcal{B}_{r,\varepsilon}=\mathcal{B}_{1,r,\varepsilon}\cup\cdots\cup\mathcal{B}_{k,r,\varepsilon}$ and use the union bound to get

$$\frac{1}{L(r)}\sum_{r\leq n\leq r+L(r)} \Lambda_{w,b}(n)\mathbf{1}_{\mathcal{B}_{r,\varepsilon}}(n) \ll_k \varepsilon + +o_w(1)\log\frac{1}{\varepsilon}+o_r(1),$$

provided that r is very large in terms of $1/\varepsilon, w$. This is the desired conclusion. □

4.5. **Simultaneous approximation of Hardy field functions.** In view of Proposition 4.4, 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 4.5 and 4.6 that can cover them adequately. We refer again to our notational conventions in Section 1 for the notation $\lll$.

Proposition 4.7. *Let $\ell\in\mathbb{N}$ and suppose $a_1,\dots,a_\ell\in\mathcal{H}$ 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*

$$1 \lll \big|a_i^{(k_i)}(t)\big|^{-\frac{1}{k_i}} \lll L(t) \lll \big|a_i^{(k_i+1)}(t)\big|^{-\frac{1}{k_i+1}}.$$

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 $\big|a^{(k+1)}(t)\big|^{-\frac{1}{k+1}}$ strongly dominates the function $\big|a^{(k)}(t)\big|^{-\frac{1}{k}}$. Then, we can pick the function $L(t)$ to be the geometric mean of these two functions to get our claim.[19]

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

[19]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).

the assumptions of Lemma 4.1, 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. For the sake of contradiction, we assume that this fails and use the lower bound from Lemma 4.1, to deduce that

$$t^{k(\varepsilon-1)} \gg a^{(k)}(t) \succ \frac{a(t)}{t^k \log^{2k} t}$$

for every $0 < \varepsilon < 1$. This, implies that $a(t) \ll t^{k\varepsilon} \log^{2k} t$ for all small ε, 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, as the last claim in our statement suggests.

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

$$\frac{\left|a^{(k+1)}(t)\right|^{-\frac{1}{k+1}}}{\left|a^{(k)}(t)\right|^{-\frac{1}{k}}} \gg t^{\delta} \tag{65}$$

for some $\delta > 0$ that will depend on k. Choose a real number $0 < q < 1$ (the value of q depends on k), such that $\left|a^{(k)}(t)\right|^{-\frac{1}{k}} \ll t^{1-q}$, which can be done as we demonstrated above. In order to establish (65), we combine the inequality $a^{(k)}(t) \gg ta^{(k+1)}(t)$ with the inequality $\left|a^{(k)}(t)\right|^{-\frac{1}{k}} \ll t^{1-q}$, which after some computations gives the desired result for $\delta = q/(k+1)$. This completes the base case.

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

$$\left|a_i^{(k_i)}(t)\right|^{-\frac{1}{k_i}} \lll L(t) \lll \left|a_i^{(k_i+1)}(t)\right|^{-\frac{1}{k_i+1}}, \quad 1 \le i \le \ell.$$

Due to Proposition 4.3, there exists a positive integer s, such that

$$\left|a_{\ell+1}^{(s)}(t)\right|^{-\frac{1}{s}} \prec L(t) \prec \left|a_{\ell+1}^{(s+1)}(t)\right|^{-\frac{1}{s+1}}. \tag{66}$$

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 $\left|a_{\ell+1}^{(s)}(t)\right|^{-\frac{1}{s}} \lll \left|a_{\ell+1}^{(s+1)}(t)\right|^{-\frac{1}{s+1}}$ holds, as we established in the base case of the induction.

If each function strongly dominates the preceding one in (66), then we are finished. Therefore, assume that $L(t)$ is not strongly dominated by the function $\left|a_{\ell+1}^{(s+1)}(t)\right|^{-\frac{1}{s+1}}$ (the other case is similar). Note that for every $1 \le i \le \ell$, we have that

$$\left|a_i^{(k_i)}(t)\right|^{-\frac{1}{k_i}} \lll \left|a_{\ell+1}^{(s+1)}(t)\right|^{-\frac{1}{s+1}}.$$

Indeed, since the function $L(t)$ strongly dominates the function $\left|a_i^{(k_i)}(t)\right|^{-\frac{1}{k_i}}$ (by the induction hypothesis) and $L(t)$ grows slower than the the function $\left|a_{\ell+1}^{(s+1)}(t)\right|^{-\frac{1}{s+1}}$, this claim follows immediately. Among the functions $a_1, \ldots, a_{\ell+1}$, we choose a function for which the growth rate of $\big|a_i^{(k_i)}(t)\big|^{-\frac{1}{k_i}}$ is maximized.[20] Assume that this happens for the index $i_0 \in \{1, \ldots, \ell+1\}$ and observe that the function $\big|a_{\ell+1}^{(s+1)}(t)\big|^{-\frac{1}{s+1}}$ strongly dominates $\big|a_{i_0}^{(k_{i_0})}(t)\big|^{-\frac{1}{k_{i_0}}}$, 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 $\big|a_{\ell+1}^{(s)}(t)\big|^{-\frac{1}{s}} \lll \big|a_{\ell+1}^{(s+1)}(t)\big|^{-\frac{1}{s+1}}$).

Define the function $\widetilde{L}(t)$ to be the geometric mean of the functions $\big|a_{i_0}^{(k_{i_0})}(t)\big|^{-\frac{1}{k_{i_0}}}$ and $\big|a_{\ell+1}^{(s+1)}(t)\big|^{-\frac{1}{s+1}}$. Observe that this function grows slower than the function $L(t)$, since it is strongly dominated by the function $\big|a_{\ell+1}^{(s+1)}(t)\big|^{-\frac{1}{s+1}}$, while the original function $L(t)$ is not. Due to its construction, we deduce that the function $\widetilde{L}(t)$ satisfies

$$\big|a_{\ell+1}^{(s)}(t)\big|^{-\frac{1}{s}} \lll \widetilde{L}(t) \lll \big|a_{\ell+1}^{(s+1)}(t)\big|^{-\frac{1}{s+1}}$$

and

$$\big|a_i^{(k_i)}(t)\big|^{-\frac{1}{k_i}} \lll \widetilde{L}(t)$$

for all $1 \leq i \leq \ell$. This is a simple consequence of the fact that $\widetilde{L}(t)$ strongly dominates the function $\big|a_{i_0}^{(k_{i_0})}(t)\big|^{-\frac{1}{k_{i_0}}}$ 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

$$\widetilde{L}(t) \prec L(t) \lll \big|a_i^{(k_i+1)}(t)\big|^{-\frac{1}{k_i+1}}$$

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 $\big|a_i^{(k_i+1)}(t)\big|^{-\frac{1}{k_i+1}}$ 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 4.7.

Proposition 4.8. *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*

$$1 \lll \big|a_i^{(k_i)}(t)\big|^{-\frac{1}{k_i}} \lll L(t) \lll \big|a_i^{(k_i+1)}(t)\big|^{-\frac{1}{k_i+1}}.$$

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 4.7 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. ☐

[20] In the case $i = \ell+1$, we are referring to the function $\big|a_i^{(s)}(t)\big|^{-\frac{1}{s}}$.

5. 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 4.4, 4.5 and 4.6. 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 similar to [44, Lemma 3.3], the only difference being the presence of the unbounded weights.

Lemma 5.1. *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

$$\lim_{n\to+\infty} p(n) = 1. \tag{67}$$

This follows exactly as in the proof of Lemma 3.3 in [44], so we omit its proof here.

Now, we show that

$$\frac{1}{R} \sum_{n=R+1}^{R+L(R)} p(n) \big(\Lambda_{w,b}(n) - 1\big) A_n = o_R(1). \tag{68}$$

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 2.8 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 2.8 (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. □

We restate here our main theorem for convenience.

Theorem 1.1. *Let ℓ,k be positive integers and, for all $1\leq i\leq\ell,\ 1\leq j\leq k$, let $a_{ij}\in\mathcal{H}$ be functions of polynomial growth such that*

(69) $$|a_{ij}(t)-q(t)|\succ\log t\ \ \textit{for every polynomial } q(t)\in\mathbb{Q}[t],$$

or

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

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

$$\lim_{w\to+\infty}\ \limsup_{N\to+\infty}\ \max_{\substack{1\le b\le W\\(b,W)=1}}\Big\|\frac{1}{N}\sum_{n=1}^{N}\ \big(\Lambda_{w,b}(n)-1\big)\prod_{j=1}^{\ell}\big(\prod_{i=1}^{k}T_i^{\lfloor a_{ij}(Wn+b)\rfloor}\big)f_j\Big\|_{L^2(\mu)}=0. \tag{71}$$

Proof. 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 $f_1,\dots,f_\ell$ 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 $q_{ij}(Q_0n+s_0)$ have integer coefficients except possibly the constant term, for all $0\le s_0\le Q_0$. These non-integer constant terms can be absorbed into the polynomial $p_{ij}(t)$. Therefore, splitting our average into the arithmetic progressions (Q_0n+s_0), it suffices to show that

$$\lim_{w\to+\infty}\ \limsup_{N\to+\infty}\ \max_{\substack{1\le b\le W\\(b,W)=1}}\Big\|\frac{1}{N}\sum_{n=1}^{N}\big(\Lambda_{w,b}(Q_0n+s_0)-1\big)\prod_{j=1}^{\ell}\big(\prod_{i=1}^{k}T_i^{\lfloor a_{ij,w,b}(Q_0n+s_0)\rfloor}\big)f_j\Big\|_{L^2(\mu)}=0$$

for all $s_0\in\{0,\dots,Q_0-1\}$. Observe that each one of the functions $a_{ij,w,b}(Q_0t+s_0)$ satisfies either (7) or (8). Since the polynomials $q_{ij,w,b}(Q_0n+s_0)$ have integer coefficients, we can rewrite the previous expression as

$$\lim_{w\to+\infty}\ \lim_{N\to+\infty}\ \max_{\substack{1\le b\le W\\(b,W)=1}}\Big\|\frac{1}{N}\sum_{n=1}^{N}\mathbf{1}_{s_0\ (Q_0)}(n)\big(\Lambda_{w,b}(n)-1\big)$$
$$\prod_{j=1}^{\ell}\big(\prod_{i=1}^{k}T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\big)f_j\Big\|_{L^2(\mu)}=0. \tag{72}$$

Step 2: Separating the iterates. Define the sets

$$S_1=\{(i,j)\in[1,k]\times[1,\ell]\colon g_{ij}(t)\ll t^\delta \text{ for all } \delta>0 \text{ and } p_{ij} \text{ is non-constant}\}, \tag{73}$$

and

$$S_2=\{(i,j)\in[1,k]\times[1,\ell]\colon g_{ij}(t)\ll t^\delta \text{ for all } \delta>0 \text{ and } p_{ij} \text{ is constant}\}, \tag{74}$$

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, (7) and (8) 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:

$$S_2'=\{(i,j)\in S_2\colon g_{ij}(t)\succ\log t\},\qquad S_2''=\{(i,j)\in S_2\colon g_{ij}(t)\prec 1\}. \tag{75}$$

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 4.8. 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 4.6 and Proposition 4.5 for these two collections respectively. Finally, observe that if $(i,j)\in S_2''$, then for n sufficiently large, we can write

$$\lfloor a_{ij}(Q_0n+s_0)\rfloor = q_{ij}(Q_0n+s_0)+\lfloor c_{ij}\rfloor + e_{ij,Q_0n+s_0},$$

where $e_{ij,Q_0n+s_0}\in\{0,-1\}$ and c_{ij} is a constant term arising from the constant (in this case) polynomial p_{ij}. The error term e_{ij,Q_0n+s_0} actually exists only if c_{ij} is an integer. In particular, we have $e_{ij,Q_0n+s_0}=0$ 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 (72), where n will be replaced by $Wn+b$.

We rewrite the limit in (72) as

$$\lim_{w\to+\infty}\ \limsup_{N\to+\infty}\ \max_{\substack{1\le b\le W\\ (b,W)=1}}\Big\|\frac{1}{N}\sum_{n=1}^{N}\mathbf{1}_{s_0\ (Q_0)}(n)\big(\Lambda_{w,b}(n)-1\big) \tag{76}$$

$$\prod_{j=1}^{\ell}\Big(\prod_{i\colon (i,j)\in S_1} T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\cdot\prod_{i\colon (i,j)\in S_2'} T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\cdot$$

$$\prod_{i\colon (i,j)\in S_2''} T_i^{q_{ij,w,b}(n)}\cdot\prod_{i\colon (i,j)\notin S_1\cup S_2'\cup S_2''} T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\Big)f_j\Big\|_{L^2(\mu)}.$$

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 4.6, 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 4.8 (thus, each one of them satisfies Proposition 4.4 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 4.5. 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

$$t^{\frac{5}{8}}\lll L(t)\lll t \tag{77}$$

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

$$\left|g_{ij}^{(k_{ij})}(t)\right|^{-\frac{1}{k_{ij}}}\lll L(t)\lll\left|g_{ij}^{(k_{ij}+1)}(t)\right|^{-\frac{1}{k_{ij}+1}}. \tag{78}$$

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 (78) also implies the inequalities

$$\left|g_{ij,w,b}^{(k_{ij})}(t)\right|^{-\frac{1}{k_{ij}}}\lll L(t)\lll\left|g_{ij,w,b}^{(k_{ij}+1)}(t)\right|^{-\frac{1}{k_{ij}+1}}. \tag{79}$$

for any fixed w,b.

For the choice of $L(t)$ that we made above, we apply Lemma 5.1 to infer that it suffices to show that

$$\lim_{w\to+\infty}\ \limsup_{R\to+\infty}\ \max_{\substack{1\leq b\leq W\\(b,W)=1}}\ \mathop{\mathbb{E}}_{1\leq r\leq R}\Big\| \mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} \mathbf{1}_{s_0\ (Q_0)}(n)\big(\Lambda_{w,b}(n)-1\big)$$
$$\prod_{j=1}^{\ell}\Big(\prod_{i\colon (i,j)\in S_1} T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\cdot \prod_{i\colon (i,j)\in S_2'} T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\cdot$$
$$\prod_{i\colon (i,j)\in S_2''} T_i^{q_{ij,w,b}(n)}\cdot \prod_{i\colon (i,j)\notin S_1\cup S_2'\cup S_2''} T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\Big) f_j\Big\|_{L^2(\mu)}=0. \tag{80}$$

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

$$\mathcal{J}_{w,b,s_0}(R) := \mathop{\mathbb{E}}_{1\leq r\leq R}\Big\| \mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} \mathbf{1}_{s_0\ (Q_0)}(n)\big(\Lambda_{w,b}(n)-1\big)$$
$$\prod_{j=1}^{\ell}\Big(\prod_{i\colon (i,j)\in S_1} T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\cdot \prod_{i\colon (i,j)\in S_2'} T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\cdot$$
$$\prod_{i\colon (i,j)\in S_2''} T_i^{q_{ij,w,b}(n)}\cdot \prod_{i\colon (i,j)\notin S_1\cup S_2'\cup S_2''} T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\Big) f_j\Big\|_{L^2(\mu)}. \tag{81}$$

We will apply Propositions 4.4, 4.5 and 4.6 to replace the iterates with polynomials (with coefficients depending on r). Due to the nature of Proposition 4.5 (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 (75), 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 4.5: for all except at most εR values of $r\in[1,R]$, we have that

$$\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor = \lfloor g_{ij,w,b}(r)+p_{ij,w,b}(r)\rfloor \text{ for all } n\in[r,r+L(r)]. \tag{82}$$

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 (79) 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 4.4 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

$$\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor=\lfloor \widetilde{p}_{ij,w,b,r}(n)\rfloor, \tag{83}$$

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

$$\frac{\big|\{n\in[r,r+L(r)]\colon \{\widetilde{p}_{ij,w,b,r}(n)\}\in[1-\delta,1)\}\big|}{L(r)}=\delta+O_A(\log^{-A} r) \tag{84}$$

for any $\delta<1$. Practically, this last condition signifies that the polynomials $\widetilde{p}_{ij,w,b,r}$ satisfy the equidistribution condition in Proposition 3.1, which we shall invoke later.

Case 3 : Finally, we deal with the case of the set S_1. Proposition 4.6 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

$$\lfloor p_{ij,w,b}(n)+g_{ij,w,b}(n)\rfloor=\lfloor p_{ij,w,b}(n)+g_{ij,w,b}(r)\rfloor. \tag{85}$$

Additionally, the set $\mathcal{B}_{w,b,r,\varepsilon}$ satisfies

$$\frac{1}{L(r)}\sum_{r\le n\le r+L(r)}\Lambda_{w,b}(n)\mathbf{1}_{\mathcal{B}_{w,b,r,\varepsilon}}(n)\ll_{k,\ell,d}\ \varepsilon+o_w(1)\log\frac{1}{\varepsilon}+o_r(1). \tag{86}$$

We emphasize that the asymptotic constant in (86) 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 (82) 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

$$\Big\|\mathop{\mathbb{E}}_{r\le n\le r+L(r)}\mathbf{1}_{s_0\ (Q_0)}(n)\big(\Lambda_{w,b}(n)-1\big)\prod_{j=1}^{\ell}\Big(\prod_{i\colon (i,j)\in S_1}T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\cdot$$
$$\prod_{i\colon (i,j)\in S_2'}T_i^{\lfloor g_{ij,w,b}(r)+p_{ij,w,b}(r)\rfloor+q_{ij,w,b}(n)}\cdot\prod_{i\colon (i,j)\in S_2''}T_i^{q_{ij,w,b}(n)}\cdot$$
$$\prod_{i\colon (i,j)\notin S_1\cup S_2'\cup S_2''}T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\Big)f_j\Big\|_{L^2(\mu)}.$$

Thus, we have replaced the iterates of the set S_2' with polynomials in the averaging parameter n.

Secondly, we use (83) 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

$$\prod_{j=1}^{\ell}\Big(\prod_{i\colon (i,j)\in S_1}T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\prod_{i\colon (i,j)\in S_2'}T_i^{\lfloor g_{ij,w,b}(r)+p_{ij,w,b}(r)\rfloor+q_{ij,w,b}(n)}\cdot$$
$$\prod_{i\colon (i,j)\in S_2''}T_i^{q_{ij,w,b}(n)}\cdot\prod_{i\colon (i,j)\notin S_1\cup S_2'\cup S_2''}T_i^{\lfloor \widetilde{p}_{ij,w,b}(n)\rfloor+q_{ij,w,b}(n)}\Big)f_j. \tag{87}$$

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 $A \geq 2$, this contribution is $o_r(1)$ and we can rewrite the average over the corresponding short interval as

$$\Big\| \mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} \mathbf{1}_{s_0\ (Q_0)}(n)\big(\Lambda_{w,b}(n)-1\big)\prod_{j=1}^{\ell}\Big(\prod_{i\colon (i,j)\in S_1} T_i^{\lfloor g_{ij,w,b}(n)+p_{ij,w,b}(n)\rfloor + q_{ij,w,b}(n)} \prod_{i\colon (i,j)\in S_2'} T_i^{\lfloor g_{ij,w,b}(r)+p_{ij,w,b}(r)\rfloor + q_{ij,w,b}(n)} \cdot \prod_{i\colon (i,j)\in S_2''} T_i^{q_{ij,w,b}(n)} \cdot \prod_{i\colon (i,j)\notin S_1\cup S_2'\cup S_2''} T_i^{\lfloor \widetilde{p}_{ij,w,b,r}(n)\rfloor + q_{ij,w,b}(n)}\Big) f_j\Big\|_{L^2(\mu)} + o_r(1). \tag{88}$$

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 (85) to rewrite (87) 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

$$\frac{1}{L(r)} \sum_{r\leq n\leq r+L(r)} \Lambda_{w,b}(n)\mathbf{1}_{\mathcal{B}_{w,b,r,\varepsilon}}(n) + \frac{1}{L(r)} \sum_{r\leq n\leq r+L(r)} \mathbf{1}_{\mathcal{B}_{w,b,r,\varepsilon}}(n). \tag{89}$$

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 (86). 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 (88) is bounded by $O_{k,\ell,d}(1)$ times

$$\Big\| \mathop{\mathbb{E}}_{r\leq n\leq r+L(r)} \mathbf{1}_{s_0(Q_0)}(n)\big(\Lambda_{w,b}(n)-1\big)\prod_{j=1}^{\ell}\Big(\prod_{i\colon (i,j)\in S_1} T_i^{\lfloor g_{ij,w,b}(r)+p_{ij,w,b}(n)+q_{ij,w,b}(n)\rfloor} \prod_{i\colon (i,j)\in S_2'} T_i^{\lfloor g_{ij,w,b}(r)+p_{ij,w,b}(r)+q_{ij,w,b}(n)\rfloor} \cdot \prod_{i\colon (i,j)\in S_2''} T_i^{\lfloor q_{ij,w,b}(n)\rfloor} \cdot \prod_{i\colon (i,j)\notin S_1\cup S_2'\cup S_2''} T_i^{\lfloor \widetilde{p}_{ij,w,b}(n)+q_{ij,w,b}(n)\rfloor}\Big) f_j\Big\|_{L^2(\mu)} + \varepsilon + o_w(1)\log\frac{1}{\varepsilon} + o_r(1). \tag{90}$$

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 3.1. 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 (84).

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 (90) is bounded by

$$\varepsilon^{-k\ell} \big\| \mathbf{1}_{s_0\ (Q_0)} \big(\Lambda_{w,b}(n) - 1\big) \big\|_{U^s(r, r+sL(r)]} + \varepsilon^{-k\ell} o_w(1) + o_\varepsilon(1)(1 + o_w(1)) + \varepsilon + o_w(1) \log \frac{1}{\varepsilon} + o_r(1). \tag{91}$$

Applying Lemma 2.3, we can bound the previous Gowers norm along the residue class $s_0\ (Q_0)$ as follows:

$$\big\| \mathbf{1}_{s_0\ (Q_0)} \big(\Lambda_{w,b}(n) - 1\big) \big\|_{U^s(r, r+sL(r)]} \leq \big\| \Lambda_{w,b}(n) - 1 \big\|_{U^s(r, r+sL(r)]}. \tag{92}$$

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 (81). 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 2.8 to conclude that this expression is $O(1) + o_r(1)$, using similar estimates as the ones used in the proof of Proposition 4.6 (see (64)). 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

$$\mathcal{J}_{w,b,s_0}(R) \ll_{d,k,\ell} \varepsilon^{-k\ell} \Big(\mathop{\mathbb{E}}_{1 \leq r \leq R} \big\| \big(\Lambda_{w,b}(n) - 1\big) \big\|_{U^s(r, r+sL(r)]} \Big) + \varepsilon^{-k\ell} o_w(1) + o_\varepsilon(1)(1 + o_w(1)) + o_R(1). \tag{93}$$

We restate (80) here. Namely, we want to show that

$$\limsup_{R\to+\infty} \max_{\substack{1\le b\le W\\(b,W)=1}} \mathcal{J}_{w,b,s_0}(R) = o_w(1).$$

Applying (93), we conclude that for a fixed w, we have

$$\limsup_{R\to+\infty} \max_{\substack{1\le b\le W\\(b,W)=1}} \mathcal{J}_{w,b,s_0}(R) \ll_{d,k,\ell} \varepsilon^{-k\ell}\Big(\lim_{R\to+\infty} \mathop{\mathbb{E}}_{1\le r\le R} \max_{\substack{1\le b\le W\\(b,W)=1}} \big\|\big(\Lambda_{w,b}(n)-1\big)\big\|_{U^s(r,r+L(r)]}\Big)+$$
$$\varepsilon^{-k\ell}o_w(1) + o_\varepsilon(1)(1+o_w(1)).$$

Due to Theorem A, we have that

$$\max_{\substack{1\le b\le W\\(b,W)=1}} \big\|\big(\Lambda_{w,b}(n)-1\big)\big\|_{U^s(r,r+L(r)]} = o_w(1)$$

for every sufficiently large r. Thus, we conclude that

$$\limsup_{R\to+\infty} \max_{\substack{1\le b\le W\\(b,W)=1}} \mathcal{J}_{w,b,s_0}(R) \ll_{d,k,\ell} \varepsilon^{-k\ell}o_w(1) + o_\varepsilon(1)(1+o_w(1)).$$

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

(94)
$$\limsup_{R\to+\infty} \lim_{N\to+\infty} \max_{\substack{1\le b\le W\\(b,W)=1}} \Big\| \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}_{s_0(Q_0)}(n)\big(\Lambda_{w,b}(n)-1\big) \prod_{j=1}^{\ell} \big(\prod_{i=1}^{k} T_i^{\lfloor a_{ij,w,b}(n)\rfloor}\big) f_j \Big\|_{L^2(\mu)}$$
$$\ll_{d,k,\ell} \varepsilon^{-k\ell}o_w(1) + o_\varepsilon(1)(1+o_w(1)),$$

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 (72) (in view of (94)) 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. □

6. Proofs of the remaining theorems

We finish the proofs of our theorems in this section.

6.1. Proof of the convergence results.

Proof of Theorem 1.2. Let $(X,\mathcal{X},\mu,T_1,\dots,T_k)$ be the system and $a_{ij}\in\mathcal{H}$ the functions in the statement. In view of Lemma 2.6, 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 $W=\prod_{p\le w, p\in\mathbb{P}} p$ as usual and let $b\in\mathbb{N}$. We define

$$B_{w,b}(N) := \frac{1}{N}\sum_{n=1}^{N} \Lambda(n)\big(\prod_{i=1}^{k} T_i^{\lfloor a_{i1}(Wn+b)\rfloor}\big)f_1\cdot\ldots\cdot\big(\prod_{i=1}^{k} T_i^{\lfloor a_{i\ell}(Wn+b)\rfloor}\big)f_\ell.$$

Let $\varepsilon>0$. Using Theorem 1.1, we can find $w_0\in\mathbb{N}$ (which yields a corresponding W_0) such that

(95)
$$\Big\| A(W_0N) - \frac{1}{\phi(W_0)} \sum_{\substack{1\le b\le W_0\\(b,W_0)=1}} B_{w_0,b}(N) \Big\|_{L^2(\mu)} = O(\varepsilon)$$

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 (95), 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 (95) 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. Let $a \in \mathcal{H}$ satisfy either (11) or (12), $k \in \mathbb{N}$, $(X, \mathcal{X}, \mu, T)$ 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 (7) or (8). 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 [10, Theorem 2.1],[21] 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 $L^2(\mu)$. aWe conclude that the two conditions of Theorem 1.2 are satisfied, which shows that the desired averages converge.

In particular, if a satisfies condition (11), we can invoke [10, 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 1.2 yields the desired conclusion. □

Proof of Theorem 1.4. We work analogously as in the proof of Theorem 1.3. The only difference is that in this case, we use [44, Theorem 1.2] 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 1.2. □

Proof of Theorem 1.5. The proof follows identically as the one of Theorem 1.4 by using [11, Theorem 2.3] instead of [44, Theorem 1.2]. □

[21] There is a slight issue 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 [10], However, the proof in [10] 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 $a \in \mathcal{H}$.

6.2. **Proof of the recurrence results.** We recall Furstenberg's Correspondence Principle for $\mathbb{Z}^d$-actions [19], for the reader's convenience.

Theorem F (Furstenberg's Correspondence Principle)**.** *Let $d \in \mathbb{N}$ and $E \subseteq \mathbb{Z}^d$. There exists a system $(X, \mathcal{X}, \mu, T_1, \ldots, T_d)$ and a set $A \in \mathcal{X}$ with $\bar{d}(E) = \mu(A)$, such that*

$$\bar{d}\big(E \cap (E + \mathbf{n}_1) \cap \cdots \cap (E - \mathbf{n}_k)\big) \geq \mu\left(A \cap \prod_{i=1}^{d} T_i^{-n_{i,1}} A \cap \cdots \cap \prod_{i=1}^{d} T_i^{-n_{i,k}} A\right),$$

for all $k \in \mathbb{N}$ and $\mathbf{n}_j = (n_{1,j}, \ldots, n_{d,j}) \in \mathbb{Z}^d$, $1 \leq j \leq k$.

In view of the correspondence principle, the corollaries in Section 1 follow easily.

Proof of Theorem 1.6. (a) We apply Theorem 1.3 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

$$\lim_{N\to+\infty} \frac{1}{\pi(N)} \sum_{p\in\mathbb{P}\colon p\leq N} T^{\lfloor a(p)\rfloor}\mathbf{1}_A \cdot \ldots \cdot T^{k\lfloor a(p)\rfloor}\mathbf{1}_A = \lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} T^{n}\mathbf{1}_A \cdot \ldots \cdot T^{kn}\mathbf{1}_A,$$

and applying Furstenberg's multiple recurrence theorem we infer that

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

which is the desired result.

(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 [29, 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

$$\liminf_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} \mu(A \cap T^{-[[cq(n)+\varepsilon(n)]]}A \cap \cdots \cap T^{-k[[cq(n)+\varepsilon(n)]]}A) \geq c_0. \tag{96}$$

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 (96) 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 1.1 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 1.8. Similarly to the proof of Theorem 1.6, we apply Theorem 1.4 for the functions $f_1 = \cdots = f_k = \mathbf{1}_A$. We deduce that

(97)
$$\lim_{N\to+\infty} \frac{1}{\pi(N)}\sum_{p\in\mathbb{P}\colon p\leq N} \mu\big(A\cap T^{-\lfloor a_1(p)\rfloor}A\cap\cdots\cap T^{-\lfloor a_k(p)\rfloor}A\big) = \int \mathbf{1}_A\cdot\big(\mathbb{E}(\mathbf{1}_A|\mathcal{I}(T))\big)^k\,d\mu.$$

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 1.10. The proof is similar to the proof of Theorem 1.8. The only distinction is made in (97), namely we have

$$\lim_{N\to+\infty} \frac{1}{\pi(N)}\sum_{p\in\mathbb{P}\colon p\leq N} \mu\big(A_0\cap T_1^{-\lfloor a_1(p)\rfloor}A_1\cap\cdots\cap T_k^{-\lfloor a_k(p)\rfloor}A_k\big) =$$
$$\int \mathbf{1}_{A_0}\cdot\mathbb{E}(\mathbf{1}_{A_1}|\mathcal{I}(T_1))\cdot\ldots\cdot\mathbb{E}(\mathbf{1}_{A_k}|\mathcal{I}(T_k))\,d\mu,$$

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, [6, Lemma 1.6] implies that

$$\int f\cdot\mathbb{E}(f|\mathcal{I}(T_1))\cdot\ldots\cdot\mathbb{E}(f|\mathcal{I}(T_k))\,d\mu \geq \Big(\int f\,d\mu\Big)^{k+1} = \mu(A)^{k+1},$$

and the conclusion follows. □

6.3. **Proof of the equidistribution results in nilmanifolds.** In this final part of this section, we offer a proof for Theorem 1.12. The main tool is the approximation of Lemma 2.12.

Proof of Theorem 1.12. 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:

$$\lim_{N\to+\infty} \frac{1}{\pi(N)} \sum_{p\in\mathbb{P}:\, p\leq N} f_1(g_1^{\lfloor a_1(p)\rfloor}x_1)\cdot\ldots\cdot f_k(g_k^{\lfloor a_k(p)\rfloor}x_k) = \int_{Y_1} f_1\, dm_{Y_1}\cdot\ldots\cdot\int_{Y_k} f_k\, dm_{Y_k},$$

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

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} \Lambda(n) f_1(g_1^{\lfloor a_1(n)\rfloor}x_1)\cdot\ldots\cdot f_k(g_k^{\lfloor a_k(n)\rfloor}x_k) = \int_{Y_1} f_1\, dm_{Y_1}\cdot\ldots\cdot\int_{Y_k} f_k\, dm_{Y_k}, \tag{98}$$

where the equivalence of the last two relations is a consequence of Lemma 2.6.

Our equidistribution assumption implies that for all $W, b\in\mathbb{N}$, we have

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} f_1(g_1^{\lfloor a_1(Wn+b)\rfloor}x_1)\cdot\ldots\cdot f_k(g_k^{\lfloor a_k(Wn+b)\rfloor}x_k) = \int_{Y_1} f_1\, dm_{Y_1}\cdot\ldots\cdot\int_{Y_k} f_k\, dm_{Y_k}. \tag{99}$$

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 $F: Y\to\mathbb{C}$ by $F(y_1,\ldots,y_k) = f_1(y_1)\cdot\ldots\cdot f_k(y_k)$ and rewrite (98) as

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} \Lambda(n) F(\widetilde{g}_1^{\lfloor a_1(n)\rfloor}\cdot\ldots\cdot \widetilde{g}_k^{\lfloor a_k(n)\rfloor}\widetilde{x}) = \int_Y F\, dm_Y, \tag{100}$$

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 (99) as

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} F(\widetilde{g}_1^{\lfloor a_1(Wn+b)\rfloor}\cdot\ldots\cdot \widetilde{g}_k^{\lfloor a_k(Wn+b\rfloor}\widetilde{x}) = \int_Y F\, dm_Y. \tag{101}$$

Therefore, we want to prove (100) under the assumption that (101) 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 2.12 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 $a_1,\dots,a_k$ satisfy the assumptions of Theorem 1.1. 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

$$\max_{\substack{1\leq b\leq W\\(b,W_0)=1}}\Big|\frac{1}{N}\sum_{n=1}^{N}\big(\Lambda_{w_0,b}(n)-1\big)\int\prod_{j=1}^{s+1}\Big(\prod_{i=1}^{k}S_i^{\ell_j\lfloor a_i(W_0n+b)\rfloor}\Big)G_j\,d\mu\Big|\leq\varepsilon \tag{102}$$

for every sufficiently large $N\in\mathbb{N}$. In addition, we use (101), 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

$$\Big|B'_{W_0,b}(N)-\int_Y F\,dm_Y\Big|\leq 2\varepsilon, \tag{103}$$

for all $1\leq b\leq W_0$ coprime to W_0.

Observe that (102) implies that for all N sufficiently large, we have

$$\Big|A'(W_0N)-\frac{1}{\phi(W_0)}\sum_{\substack{1\leq b\leq W_0\\(b,W_0)=1}}B'_{W_0,b}(N)\Big|\leq 2\varepsilon,$$

and we can combine this with (103) to conclude that

$$\Big|A'(W_0N)-\int_Y F\,dm_Y\Big|\leq 4\varepsilon$$

for all N sufficiently large. Since $|A'(N)-A(N)|\leq\varepsilon(1+o_N(1))$, we finally arrive at the inequality

$$\Big|A(W_0N)-\int_Y F\,dm_Y\Big|\leq 6\varepsilon,$$

for all large enough $N\in\mathbb{N}$. Since $|A(W_0N)-A(W_0N+b)|=o_N(1)$ for all $1\leq b\leq W$, we conclude that

$$\Big|A(N)-\int_Y F\,dm_Y\Big|\leq 7\varepsilon,$$

for all sufficiently large $N\in\mathbb{N}$. Sending $\varepsilon\to 0$, we deduce (100), which is what we wanted to show. □

Proof of Proposition Corollary 1.13. The result follows readily from Theorem 1.12. The first hypothesis of the criterion is satisfied, since each of the functions $a_i(t)$ satisfies (16), while condition (b) follows from [43, Theorem 1.1] and our assumption that $a_i(Wt+b)$ belongs to $\mathcal{H}$. □

7. More general iterates

In this last section of the article, 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 4.4, 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 1.4, 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. Nonetheless, this is not feasible in the case of one function, such as Theorem 1.3, which leads to Szemerédi-type results.

We have the following proposition.

Proposition 7.1. *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*[22]

$$t^{5/8} \ll \left|a^{(k)}(t)\right|^{-\frac{1}{k}} \lll \left|a^{(k+1)}(t)\right|^{-\frac{1}{k+1}} \ll t.$$

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

$$\lim_{w\to+\infty} \limsup_{N\to+\infty} \max_{\substack{1\leq b\leq W\\ (b,W)=1}} \Big\| \frac{1}{N}\sum_{n=1}^{N} \big(\Lambda_{w,b}(n)-1\big) T_1^{\lfloor a(Wn+b)\rfloor} f_1 \cdot \ldots \cdot T_\ell^{\lfloor a(Wn+b)\rfloor} f_\ell \Big\|_{L^2(\mu)} = 0.$$

We remark that any improvement in the parameter 5/8 in Theorem A will also lower the term $t^{5/8}$ on the leftmost part of the growth inequalities accordingly.

Sketch of the proof of Proposition 7.1. 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

$$\limsup_{N\to+\infty} \max_{\substack{1\leq b\leq W\\ (b,W)=1}} \Big\| \frac{1}{N}\sum_{n=1}^{N} \big(\Lambda_{w,b}(n)-1\big) T_1^{\lfloor g(Wn+b)\rfloor} f_1 \cdot \ldots \cdot T_\ell^{\lfloor g(Wn+b)\rfloor} f_\ell \Big\|_{L^2(\mu)} = o_w(1).$$

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 [10, 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 (68). [23]

Using Proposition 4.4 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 4.4, which implies that Proposition 3.1 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 1.1, so we omit the rest of the details. □

[22] See the subsection with the notational conventions in Section 1 for the notation $\lll$.

[23] In particular, this case is much simpler than the method used to establish Theorem 1.1, 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 $L(t) \prec t$.

An application of the previous comparison is for the class of *tempered* functions, which we define promptly.

Definition 7.2. *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

$$g_1(t) = t^{1/25}(100 + \sin\log t)^3,\ g_2(t) = t^{1/25},\ g_3(t) = t^{17/2}(2 + \cos\sqrt{\log t}). \tag{104}$$

We have that g_1 and g_2 are Fejér, g_3 is tempered of degree 8 (which is not Hardy, see [1]). 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 [1]), so that, under the obvious modification of Definition 2.1, 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.[24]

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 (104) 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 (104) (which was first studied in [8]) 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.

In order to sidestep the aforementioned problematic cases, we restrict our study to the following subclass of tempered functions (see also [1], [31]).

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 (104)).

Notice that while the class of Fejér functions contain sub-fractional functions, $\mathcal{T}_0$ does not as, according to [7, 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 7.1.

Lemma 7.3. *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, $t^c \prec \left|g^{(k)}(t)\right|^{-\frac{1}{k}}$.

[24] For Fejér functions this is a classical result due to Fejér (for a proof see [32]). The general case follows inductively by van der Corput's difference theorem.

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 [7]). □

Using Proposition 7.1 and [10, Theorem 2.2] we get the following result. More precisely, we use the fact here that [10, 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 [10, Subsection 2.1.5]). We omit its proof as it is identical to the one of Theorem 1.3.

Theorem 7.4. *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*

$$\lim_{N\to+\infty}\frac{1}{\pi(N)}\sum_{p\in\mathbb{P}\colon p\leq N}T^{\lfloor g(p)\rfloor}f_1\cdot\ldots\cdot T^{k\lfloor g(p)\rfloor}f_k=\lim_{N\to+\infty}\frac{1}{N}\sum_{n=1}^{N}T^nf_1\cdot\ldots\cdot T^{kn}f_k, \tag{105}$$

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

As in the Hardy field case, we have the corresponding recurrence result.

Theorem 7.5. *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*

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

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 7.6. *Let $g\in\mathcal{T}$. For any set $E\subseteq\mathbb{N}$ of positive upper density, and $k\in\mathbb{N}$, we have*

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

Comment. In Theorem 7.4, and, thus, in Theorem 7.5 and Corollary 7.6, 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 $a\neq 0$.

As we saw, the comparison method provides results along primes through the corresponding results for averages along $\mathbb{N}$, though in the case of tempered functions, we do not have a comparison result of the same strength as Theorem 1.1. Nonetheless, it is expected that convergence results along $\mathbb{N}$ for iterates which are comprised of multiple tempered functions (or even combinations of tempered and Hardy field functions) can be transferred to the prime setting. Even in the case of averages along $\mathbb{N}$, the convergence results are still not established under the most general expected assumptions. For a single function and commuting transformations, a result in this direction was proven in [7]. We note that [7, Theorem 6.1] reflects the complexity of the assumptions we have to impose on the growth rates of functions to deduce such results. This analysis is beyond the scope of this paper.

References

[1] 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.

[2] 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.

[3] 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.

[4] M. Boshernitzan. Uniform distribution and Hardy fields. *Journal d'Analyse Mathématique*, **62**:225–240, 1994.

[5] 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.

[6] Q. Chu. Multiple recurrence for two commuting transformations. *Ergodic Theory and Dynamical Systems*, **31**(3):771–792, 2011.

[7] S. Donoso, A. Koutsogiannis, and W. Sun. Joint ergodicity for functions of polynomial growth. Preprint 2023, arXiv:2301.06911.

[8] S. Donoso, A. Koutsogiannis, and W. Sun. Pointwise multiple averages for sublinear functions. *Ergodic Theory and Dynamical Systems*, **40**:1594–1618, 2020.

[9] N. Frantzikinakis. Equidistribution of sparse sequences on nilmanifolds. *Journal d'Analyse Mathématique*, **109**(1):353–395, 2009.

[10] N. Frantzikinakis. Multiple recurrence and convergence for Hardy sequences of polynomial growth. *Journal d'Analyse Mathématique*, **112**(1):79–135, 2010.

[11] N. Frantzikinakis. A multidimensional Szemerédi theorem for Hardy sequences of different growth. *Transactions of the American Mathematical Society*, **367**:5653–5692, 2012.

[12] N. Frantzikinakis. Some open problems on multiple ergodic averages. *Bulletin of the Hellenic Mathematical Society*, **60**:41–90, 2016.

[13] N. Frantzikinakis. Joint ergodicity of fractional powers of primes. *Forum of Mathematics, Sigma*, **10**:e30, 2022.

[14] N. Frantzikinakis. Joint ergodicity of sequences. *Advances in Mathematics*, **417**:108918, 2023.

[15] N. Frantzikinakis and B. Host. Weighted multiple ergodic averages and correlation sequences. *Ergodic Theory and Dynamical Systems*, **38**:81–142, 2015.

[16] 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.

[17] N. Frantzikinakis, B. Host, and B. Kra. The polynomial multidimensional szemerédi theorem along shifted primes. *Israel Journal of Mathematics*, **194**:331–348, 2010.

[18] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *Journal d'Analyse Mathématique*, **31**(1):204–256, 1977.

[19] H. Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory.* Princeton University Press, 1981.

[20] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics*, **167**(2):481–547, 2008.

[21] B. Green and T. Tao. Linear equations in primes. *Annals of Mathematics*, **171**:1753–1850, 2010.

[22] B. Green and T. Tao. The Möbius function is strongly orthogonal to nilsequences. *Annals of Mathematics*, **175**(2):541–566, 2012.

[23] B. Green and T. Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Annals of Mathematics*, **175**:465–540, 2012.

[24] 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.

[25] G. H. Hardy. Properties of logarithmico-exponential functions. *Proceedings of the London Mathematical Society*, **s2-10**(1):54–90, 1912.

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

[27] B. Host and B. Kra. *Nilpotent structures in ergodic theory.* Mathematical Surveys and Monographs, Volume 236, American Mathematical Society, Providence, RI, 2018.

[28] M. N. Huxley. On the difference between consecutive primes. *Inventiones Mathematicae*, **15**:164–170, 1971.

[29] A. Koutsogiannis. Closest integer polynomial multiple recurrence along shifted primes. *Ergodic Theory and Dynamical Systems*, **38**(2):666–685, 2018.

[30] A. Koutsogiannis. Integer part polynomial correlation sequences. *Ergodic Theory and Dynamical Systems*, **38**(4):1525–1542, 2018.
[31] A. Koutsogiannis. Multiple ergodic averages for tempered functions. *Discrete and Continuous Dynamical Systems*, **41**(3):1177–1205, 2021.
[32] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences.* Pure and Applied Mathematics. Wiley-Interscience, New York-London-Sydney, 1974.
[33] 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.
[34] D. Leitmann. On the uniform distribution of some sequences. *Journal of the London Mathematical Society*, **14**:430–432, 1976.
[35] 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.
[36] 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.
[37] H. L. Montgomery and R. C. Vaughan. The large sieve. *Mathematika*, **20**:119–134, 1973.
[38] 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.
[39] F. Richter. Uniform distribution in nilmanifolds along functions from a Hardy field. *Journal d'Analyse Mathématique*, **149**:421–483, 2023.
[40] A. Sárközy. On difference sets of sequences of integers. III. *Acta Mathematica Hungarica*, **31**:355–386, 1978.
[41] I. Stux. On the uniform distribution of prime powers. *Communications on Pure and Applied Mathematics*, **27**:729–740, 1974.
[42] W. Sun. Multiple recurrence and convergence for certain averages along shifted primes. *Ergodic Theory and Dynamical Systems*, **35**(5):1592–1609, 2015.
[43] K. Tsinas. Pointwise convergence in nilmanifolds along smooth functions of polynomial growth. To appear in Ergodic Theory and Dynamical Systems, arXiv:2203.11609.
[44] K. Tsinas. Joint ergodicity of Hardy field sequences. *Transactions of the American Mathematical Society*, **376**:3191–3263, 2023.
[45] M. Wierdl. Pointwise ergodic theorem along the prime number. *Israel Journal of Mathematics*, **64**:315–336, 1988.
[46] D. Wolke. Zur gleichverteilung einiger zahlenfolgen. *Mathematische Zeitschrift*, **142**:181–184, 1975.
[47] T. Wooley and T. Ziegler. Multiple recurrence and convergence along the primes. *American Journal of Mathematics*, **134**, 2010.

(Andreas Koutsogiannis) Department of Mathematics, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece
Email address: `akoutsogiannis@math.auth.gr`

(Konstantinos Tsinas) Department of Mathematics and Applied Mathematics, University of Crete, Voutes University Campus, Heraklion 70013, Greece
Email address: `kon.tsinas@gmail.com`