Ergod. Th. & Dynam. Sys. (2020), **40**, 1594–1618
doi:10.1017/etds.2018.118

Pointwise multiple averages for sublinear functions

SEBASTIÁN DONOSO†, ANDREAS KOUTSOGIANNIS‡ and WENBO SUN‡

† *Instituto de Ciencias de la Ingeniería,*
Universidad de O'Higgins,
Av. Libertador Bernardo O'Higgins 611, Rancagua, 2841959, Chile
(e-mail: sebastian.donoso@uoh.cl)
‡ *Department of Mathematics,*
The Ohio State University,
231 West 18th Avenue, Columbus, OH 43210-1174, USA
(e-mail: koutsogiannis.1@osu.edu, sun.1991@osu.edu)

(*Received* 5 *November* 2017 *and accepted in revised form* 6 *September* 2018)

Abstract. For any measure-preserving system $(X, \mathcal{B}, \mu, T_1, \ldots, T_d)$ with no commutativity assumptions on the transformations T_i, $1 \le i \le d$, we study the pointwise convergence of multiple ergodic averages with iterates of different growth coming from a large class of sublinear functions. This class properly contains important subclasses of Hardy field functions of order zero and of Fejér functions, i.e., tempered functions of order zero. We show that the convergence of the single average, via an invariant property, implies the convergence of the multiple one. We also provide examples of sublinear functions which are, in general, bad for convergence on arbitrary systems, but good for uniquely ergodic systems. The case where the fastest function is linear is addressed as well, and we provide, in all the cases, an explicit formula of the limit function.

Key words: pointwise convergence, multiple averages, sublinear functions, Fejér functions, Hardy functions
2010 Mathematics Subject Classification: 37A05 (Primary); 37A30, 28A99 (Secondary)

1. *Introduction and main results*

The study of the limiting behavior, in $L^2(\mu)$ or pointwise, as $N \to \infty$, of *multiple ergodic averages* of the form

$$\frac{1}{N}\sum_{n=0}^{N-1} T_1^{a_1(n)} f_1 \cdots T_d^{a_d(n)} f_d, \tag{1}$$

where $T_1, \ldots, T_d : X \to X$ are invertible (usually commuting) measure-preserving transformations acting on a probability space $(X, \mathcal{B}, \mu)$; $f_1, \ldots, f_d \in L^\infty(\mu)$ and $a_i(\cdot)$ are suitable integer valued sequences for all $1 \le i \le d$†, is a central problem in ergodic theory. With partial knowledge of the limiting behavior of (1), for the case $T_i = T$ and $a_i(n) = in$, Furstenberg provided a purely ergodic theoretical proof of Szemerédi's theorem [**22**], i.e., every subset of $\mathbb{N}$ with positive upper density contains arbitrarily long arithmetic progressions. In recent years, motivated by the work of Furstenberg, fruitful progress has been made towards the study of the existence and also of the exact value of the $L^2(\mu)$ limit of (1) for various classes of sequences a_i. For the existence of the limit, we refer the readers to [**5**, **25**, **26**, **29**–**32**]. As for the explicit expression of the limit, the first result is von Neumann's mean ergodic theorem which says that, for $d = 1$ and $a_1(n) = n$, the limit of (1) is $\mathbb{E}(f_1 \mid \mathcal{I}(T_1))$, where $\mathcal{I}(T)$ denotes the σ-algebra of T-invariant sets and $\mathbb{E}(f \mid \mathcal{I}(T))$ is the *conditional expectation* of f with respect to $\mathcal{I}(T)$. The classes of integer polynomials, integer parts of polynomials, Hardy field (see definition in §2) and, more generally, tempered classes of functions are also studied in [**7**, **9**, **20**] for a single T and in [**15**, **21**, **29**] for commuting T_i.

Following the work of Gowers [**23**], the averages along cubes have enjoyed a key role in establishing convergence results, as illustrated by several papers by Host and Kra [**26**] and Bergelson [**8**], among others. For the sake of completeness, we define this notion that lies on the background of our study. Let $(X, \mathcal{B}, \mu)$ be a probability space and let T_ϵ be measure-preserving transformations on X for all $\epsilon \in \{0, 1\}^d$. The *cubic averages* are expressions of the form

$$\frac{1}{N^d} \sum_{n_1,\ldots,n_d=0}^{N-1} \prod_{\epsilon=(\epsilon_1,\ldots,\epsilon_d)\in\{0,1\}^d} T_\epsilon^{n_1\epsilon_1+\cdots+n_d\epsilon_d} f_\epsilon, \tag{2}$$

where $f_\epsilon \in L^\infty(\mu)$ for all $\epsilon \in \{0, 1\}^d$. One of the nice properties of (2) is that the pointwise convergence for such averages holds (and hence, its $L^2(\mu)$ as well) for not necessarily commuting transformations T_ϵ, $\epsilon \in \{0, 1\}^d$. This property was first discovered by Assani in the papers [**3**, **4**] and it was later extended by Chu and Frantzikinakis [**16**]. We refer the interested reader to the papers [**8**, **26**] for $L^2(\mu)$ convergence/recurrence of cubic averages for a single transformation; the papers [**5**, **14**] for $L^2(\mu)$ convergence for commuting transformations; the papers [**3**, **28**] for pointwise convergence of cubic averages for a single transformation; the papers [**17**, **19**] for the variant of cubic averages that arise from the van der Corput trick; and, finally, the papers [**3**, **4**, **16**] for pointwise convergence of cubic averages for non-commuting transformations. At this point, we should mention the important fact that the pointwise convergence along the cubes for a single transformation was obtained first by Assani (in [**3**]) and that this result is the first complete pointwise convergence result in the theory of non-conventional ergodic averages. Hence, motivated by the fruitful results on cubic averages, it is natural to ask: Does the limit of (1), as $N \to \infty$, exist in the pointwise sense? Do we necessarily have to assume that the transformations $T_1, \ldots, T_d$ commute with each other?

† For a measurable function f and a transformation $T : X \to X$, Tf denotes the composition $f \circ T$.

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

The results on the existence and explicit expression of the pointwise limit of (1) are not as abundant as that of the $L^2(\mu)$ limit, even for the case when $T_1, \ldots, T_d$ are commuting transformations. In fact, even the $d = 1$ case is not completely understood (for some results, see [**11**, **12**]). For $d = 2$ and commuting transformations T_1 and T_2, Assani gave sufficient conditions so that (1) converges pointwisely almost everywhere for all $f, g \in L^2(\mu)$ (the argument is a nice application of the classical van der Corput trick; see [**2**, Proposition 4]). For $d = 2$, Bourgain showed (in [**13**]) that the pointwise limit of (1) exists when $T_1 = T_2$ and $a_1(n) = an$, $a_2(n) = bn$ for $a, b \in \mathbb{Z}$. Recently, Huang, Shao and Ye [**27**] showed the existence of the pointwise limit of (1) for $T_i = T$, $a_i(n) = in$ under the assumption that T is a distal transformation (see also [**1**, **24**] for some particular weakly mixing systems). This result was extended in [**18**] for two commuting transformations generating a distal action and in [**19**] for an arbitrary number of commuting transformations (also for a distal system and linear iterates).

When $T_1, \ldots, T_d$ are not necessarily commuting, in order to expect (1) to have a nice behavior (in either the $L^2(\mu)$ or the pointwise sense), one needs to impose additional *distinctness* conditions on the a_i. For the case $d = 2$ and $a_1(n) = a_2(n) = n$, counterexamples where (1) does not converge in $L^2(\mu)$ were given by Berend (see [**6**, Example 7.1], where (1) does not even converge weakly) and by Bergelson and Leibman (see [**10**, §4], where T_1, T_2 generate a solvable group).

It should be mentioned that there are also positive results in the pointwise setting for $d = 2$ and $a_1(n) = a_2(n) = n$. Assani (in [**2**, Theorem 6]) provided a sufficient condition so that (1) converges pointwisely (almost everywhere for all $f, g \in L^2(\mu)$) when we do not assume any commutativity of T_1 and T_2.

On the other hand, arguably, the only known multiple convergence result for non-commuting transformations for a general system is due to Frantzikinakis. In [**20**, Theorem 2.7], under no commutativity assumption on T_i, for the integer part (denoted by $[\cdot]$) of the functions a_i in $\mathcal{LE}$†, with $x^{\varepsilon} \prec a_d \prec \cdots \prec a_1 \prec x$‡, for some $\varepsilon > 0$, he showed that the limit of (1) in $L^2(\mu)$ is equal to $\mathbb{E}(f_1 \mid \mathcal{I}(T_1)) \cdots \mathbb{E}(f_d \mid \mathcal{I}(T_d))$.

In this paper, we study the pointwise convergence of (1) for the integer part of sequences of functions of different growth rate which are sublinear (actually, the case where the fastest function is linear is addressed as well), with no commutativity assumption on the transformations. More specifically, we obtain a pointwise version of [**20**, Theorem 2.7], mentioned above, to the wider class of sublinear functions $\mathcal{S}^*$ (see the next subsection or the Appendix for notation).

THEOREM 1.1. *Let $d \in \mathbb{N}$ and (X_i, μ_i, T_i), $1 \le i \le d$ be measure-preserving systems. Let $a_i \in \mathcal{S}^*$, $1 \le i \le d$ with $a_d \prec \cdots \prec a_1$ and $a_d' \prec \cdots \prec a_1'$, and ν be any coupling of the spaces (X_i, μ_i). Then, for all $f_i \in L^\infty(\mu_i)$, $1 \le i \le d$, the averages*

$$\frac{1}{N}\sum_{n=0}^{N-1} T_1^{[a_1(n)]} f_1(x_1) \cdots T_d^{[a_d(n)]} f_d(x_d)$$

† a is a *logarithmico-exponential Hardy field function* if it belongs to a Hardy field of real-valued functions and it is defined on some $(c, +\infty)$, $c \ge 0$, by a finite combination of symbols $+, -, \times, \div, \sqrt[n]{\cdot}$, exp, log acting on the real variable x and on real constants (for more on Hardy field functions and, in particular, for logarithmico-exponential ones, check, for example, [**20**, **21**]).
‡ For two functions a, b we write $a(x) \prec b(x)$, or just $a \prec b$ if $a(x)/b(x) \to 0$ as $x \to \infty$.

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

converge as $N \to \infty$ *for* ν*-almost every (a.e.)* $(x_1, \ldots, x_d) \in X_1 \times \cdots \times X_d$ *to*

$$\mathbb{E}(f_1 \mid \mathcal{I}(T_1))(x_1) \cdots \mathbb{E}(f_d \mid \mathcal{I}(T_d))(x_d).$$

In particular, if $(X_i, \mu_i) = (X, \mu)$, $1 \le i \le d$, *for* ν *the diagonal coupling, we get that*

$$\frac{1}{N} \sum_{n=0}^{N-1} T_1^{[a_1(n)]} f_1(x) \cdots T_d^{[a_d(n)]} f_d(x)$$

converge as $N \to \infty$ *for* μ*-a.e.* $x \in X$ *to*

$$\mathbb{E}(f_1 \mid \mathcal{I}(T_1))(x) \cdots \mathbb{E}(f_d \mid \mathcal{I}(T_d))(x).$$

Remark. Let $\mathcal{LE}_\varepsilon$ denote the set of logarithmico-exponential Hardy field functions a satisfying the growth condition $x^\varepsilon \prec a(x) \prec x$ for some $\varepsilon > 0$. By a variation of the argument in [**20**, Proposition 6.4], one can obtain a different proof of Theorem 1.1 for the special case where $a_i \in \mathcal{LE}_\varepsilon$, $1 \le i \le d$. The idea of [**20**, Proposition 6.4] is to convert the multiple averages for sublinear functions of different growth in $\mathcal{LE}_\varepsilon$, via a change of variable, to an average of the same form but with a linear (equal to x) fastest function. Our method, which is applicable to a larger class of functions, has a different philosophy that focuses instead on the invariance property of the averages under the transformations $T_1 \times \mathrm{id} \times \cdots \times \mathrm{id}$, $\mathrm{id} \times T_2 \times \cdots \times \mathrm{id}, \ldots, \mathrm{id} \times \cdots \times \mathrm{id} \times T_d$, via which we deduce the limit of the expressions of interest. Another advantage of this method is that it can also be used to show that there are certain sublinear functions for which, even though the pointwise convergence might, in general, fail, it holds for all the uniquely ergodic systems (see §1.2 for details).

It is worth noting that a result similar to Theorem 1.1 holds when we replace a_1 by a linear function, i.e., a polynomial of degree one. More specifically, we have the following theorem.

THEOREM 1.2. *Let* $d \in \mathbb{N}$, (X_i, μ_i, T_i), $1 \le i \le d$ *be measure-preserving systems, let* a_1 *be a linear function,* $a_i \in \mathcal{S}^*$, $2 \le i \le d$ *with* $a_d \prec \cdots \prec a_1$ *and* $a'_d \prec \cdots \prec a'_1$, *and let* ν *be any coupling of the spaces* (X_i, μ_i). *Then, for all* $f_i \in L^\infty(\mu_i)$, $1 \le i \le d$, *the averages*

$$\frac{1}{N} \sum_{n=0}^{N-1} T_1^{[a_1(n)]} f_1(x_1) T_2^{[a_2(n)]} f_2(x_2) \cdots T_d^{[a_d(n)]} f_d(x_d)$$

converge as $N \to \infty$ *for* ν*-a.e.* $\vec{x} = (x_1, \ldots, x_d) \in X_1 \times \cdots \times X_d$.

In particular, if $a_1(n) = kn + \ell$, $k = p/q$, $p, q \in \mathbb{Z}\backslash\{0\}$, *then the limit is equal to*

$$\frac{1}{q} \sum_{j=0}^{q-1} \mathbb{E}(T_1^{[pj/q+\ell]} f_1 \mid \mathcal{I}(T_1^p))(x_1) \mathbb{E}(f_2 \mid \mathcal{I}(T_2))(x_2) \cdots \mathbb{E}(f_d \mid \mathcal{I}(T_d))(x_d),$$

while if $a_1(n) = \gamma n + \ell$, $\gamma \in \mathbb{R}\backslash\mathbb{Q}$, *then the limit is equal to*

$$F(x_1) \mathbb{E}(f_2 \mid \mathcal{I}(T_2))(x_2) \cdots \mathbb{E}(f_d \mid \mathcal{I}(T_d))(x_d),$$

where

$$F(x) = \sum_{m \in \mathbb{Z}} \exp\left(2\pi i \frac{m\ell}{\gamma}\right) \cdot \frac{\exp(-2\pi i (m/\gamma)) - 1}{-2\pi i (m/\gamma)} \mathbb{E}(f_1 \mid \mathcal{I}_{\gamma,m}(T))(x)$$

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

and $\mathcal{I}_{\gamma,m}(T)$ is the sub-σ-algebra generated by the eigenspace of T with eigenvalue $-m/\gamma$.

Via Theorems 1.1 and 1.2, applied to the diagonal coupling, we immediately get the following result on sequences of different growth rates of the form $(n^c)_n$, $0 < c \leq 1$.

COROLLARY 1.3. *Let $d \in \mathbb{N}$ and $(X, \mu, T_1, \ldots, T_d)$ be a measure-preserving system. For all $0 < c_d < \cdots < c_1 \leq 1$ and $f_1, \ldots, f_d \in L^\infty(\mu)$, the averages*

$$\frac{1}{N}\sum_{n=0}^{N-1} T_1^{[n^{c_1}]} f_1(x) \cdots T_d^{[n^{c_d}]} f_d(x) \tag{3}$$

converge as $N \to \infty$ for μ-a.e. $x \in X$ to

$$\mathbb{E}(f_1 \mid \mathcal{I}(T_1))(x) \cdots \mathbb{E}(f_d \mid \mathcal{I}(T_d))(x).$$

This answers [**20**, Problem 6] for the special case where $0 < c_d < \cdots < c_1 \leq 1$†.

1.1. *Single convergence implies multiple convergence.* The philosophy of this article is that, for a specific nice and wide class of sublinear functions, we have that 'single convergence implies the multiple one'. More specifically, assuming no commutativity on the transformations T_i, $1 \leq i \leq d$, we show, in Theorem 1.1, that averages as in (1) with integer parts of functions of different growth rate from the aforementioned class converge pointwisely and the limit is the expected one, i.e., the product of conditional expectations, using the fact that the single average converges. The same method also extends to the case when the fastest growing function is linear and we get Theorem 1.2. Our arguments throughout the article are elementary and have a soft touch of ergodic theory.

We introduce some additional notation. For $1 \leq i \leq d$, let (X_i, μ_i, T_i) be measure-preserving systems (we also assume that each (X_i, T_i) is a topological dynamical system) and let $\mu_i = \int \mu_{[T_i],x}\, d\mu_i(x)$ be the disintegration of μ_i over its factor $\mathcal{I}(T_i)$ (i.e., the *ergodic decomposition*). For $\vec{x} = (x_1, \ldots, x_d)$, let $\mu_{[T_1,\ldots,T_d],\vec{x}}$ be the measure on $X_1 \times \cdots \times X_d$ defined by

$$\int_{X_1\times\cdots\times X_d} f_1 \otimes \cdots \otimes f_d \, d\mu_{[T_1,\ldots,T_d],\vec{x}} = \mathbb{E}(f_1 \mid \mathcal{I}(T_1))(x_1) \cdots \mathbb{E}(f_d \mid \mathcal{I}(T_d))(x_d)$$

for all $f_1, \ldots, f_d \in L^\infty(\mu)$. It is easy to see that $\mu_{[T_1,\ldots,T_d],\vec{x}} = \bigotimes_{i=1}^d \mu_{[T_i],x_i}$. Let also

$$\lambda_{N,\vec{x}} := \frac{1}{N}\sum_{n=0}^{N-1} (T_1^{[a_1(n)]} \times \cdots \times T_d^{[a_d(n)]})\delta_{\vec{x}}, \tag{4}$$

where $\vec{x} \in X_1 \times \cdots \times X_d$ and $\delta_{\vec{x}}$ denotes the Dirac measure at $\vec{x}$.

Denoting with $\mathbb{R}^+$ a set of the form $(c, +\infty)$ for some $c \geq 0$, we define the class

$$S := \left\{ a \in \mathcal{C}^3(\mathbb{R}^+) \;\middle|\; a, \frac{1}{a'} \in \mathcal{SL} \text{ and } a^{-1} \in M_1 \cap D_0 \cap D_1 \cap (D_2 \cup M_2) \right\},$$

† Another possible approach to Corollary 1.3 is to use a variation of [**20**, Proposition 6.4] to convert (3), via a change of variables, to an average of the same form with $c_1 = 1$. We omit the details.

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

where, for $k \in \mathbb{N} \cup \{0\}$,

$$D_k := \left\{ a : \limsup_{x \to \operatorname{sgn}(a^{-1}) \cdot \infty} \sup_{h \in [-1,1]} \left| \frac{a^{(k+1)}(x+h)}{a^{(k)}(x)} \right| < \infty \right\},$$

$$M_k := \{a : a^{(k)} \text{ is eventually monotone}\},$$

sgn is the *sign* function† and

$$\mathcal{SL} = \{a : a(x) \prec x\}$$

is the set of *sublinear functions* (recall that $a \prec b$ means $a(x)/b(x) \to 0$ as $x \to \infty$).

Note that the lim sup that appears in the definition of D_k can, in general, be any $\alpha \in [0, \infty]$. Indeed, for $k = 0$, the function $a(x) = \log x$ gives $\alpha = 0$; to get a specific $\alpha > 0$, pick β with $\beta \exp(\beta) = \alpha$ and let $a(x) = \exp(\beta x)$; while to get $\alpha = \infty$, pick $a(x) = \exp(x^2)$.

Note also that every function $a \in \mathcal{S}$ satisfies $\log x \prec a(x)$ and $a'(x) \to 0$ as $x \to \infty$. Indeed, since $a^{-1} \in M_1$, we have that a' has eventually constant sign, and hence integrating the relation $x|a'(x)| \geq M$ (which holds eventually for $M > 0$ since $1/a' \in \mathcal{SL}$) we get $\log x \ll |a(x)|$. Using again that a, $1/a' \in \mathcal{SL}$, we have the claim since

$$\lim_{x \to \infty} \frac{\log x}{a(x)} = \lim_{x \to \infty} \frac{1/a'(x)}{x} = 0 \quad \text{and} \quad \lim_{x \to \infty} a'(x) = \lim_{x \to \infty} \frac{a(x)}{x} = 0.$$

Let $\mathcal{S}^* \subseteq \mathcal{S}$ denote the subclass of functions where $\lim_{N \to \infty} (1/N) \sum_{n=0}^{N-1} f(T^{[a(n)]}x)$ exists pointwisely (almost everywhere) for every measure-preserving system (X, μ, T) and every bounded measurable function f (we will see in §4 that, in this case, the pointwise limit is $\mathbb{E}(f \mid \mathcal{I}(T))$). We also stress the fact that $\mathcal{S}^*$ is a strict subset of $\mathcal{S}$ (see §1.2).

The following result via a density argument will lead us to the proof of Theorem 1.1.

THEOREM 1.4. *Let $d \in \mathbb{N}$ and (X_i, μ_i, T_i), $1 \leq i \leq d$ be measure-preserving systems, $a_i \in \mathcal{S}^*$, $1 \leq i \leq d$ with $a_d \prec \cdots \prec a_1$ and $a'_d \prec \cdots \prec a'_1$, and let ν be any coupling of the spaces (X_i, μ_i). Then, for ν-a.e. $\vec{x} \in X_1 \times \cdots \times X_d$, we have that $\lambda_{N,\vec{x}}$ converges to $\mu_{[T_1,\ldots,T_d],\vec{x}}$ as $N \to \infty$.*

Remark. In §2, we show that $\mathcal{S}$ contains functions a which belong to some Hardy field and satisfy $x^{\varepsilon} \prec a(x) \prec x$ for some $\varepsilon > 0$. So, by [**11**, Theorem 3.4], we actually have that each such function a is in $\mathcal{S}^*$ (with convergence to the expected limit, i.e., the conditional expectation $\mathbb{E}(f \mid \mathcal{I}(T))$) while slow Hardy field functions (as $1 \prec a(x) \prec \log x \exp((\log(\log x))^m)$ for some $0 \leq m < 1$) do not belong to $\mathcal{S}^*$ (see [**11**, Theorem 3.6]). However, even though Theorem 1.4, in general, might fail (take, for example, $a_i \in \mathcal{S} \backslash \mathcal{S}^*$ for some $1 \leq i \leq d$), we have its validity for uniquely ergodic systems and continuous functions on them, since, in this setting, the single convergence holds not only for functions in $\mathcal{S}^*$ but for all functions in $\mathcal{S}$ (see Theorem 1.5).

† The notation $\operatorname{sgn}(a^{-1}) \cdot \infty$ denotes ∞ if a^{-1} is eventually positive and $-\infty$ if a^{-1} is eventually negative.

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

1.2. *Pointwise averages on uniquely ergodic systems.* A topological system (X, T) is *uniquely ergodic* if there is a unique Borel probability measure which is T-invariant. Via the following result for single convergence, which we show in §5, under the unique ergodicity assumption of the system, we extend Theorem 1.4 (to Theorem 1.6 below).

THEOREM 1.5. *Let (X, T) be a uniquely ergodic system with unique T-invariant measure μ and $a \in \mathcal{S}$. Then, for any continuous function f on X, for every $x \in X$, we have that*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T^{[a(n)]} f(x) = \int_X f \, d\mu.$$

As we mentioned before, this result has been studied for general systems in [**11**, **12**] along functions which belong to a smaller class of functions than $\mathcal{S}$ (see Theorem 4.1), and, in general, it might fail for $a \in \mathcal{S}$. One can show (after some elementary calculations) that

$$a(x) = \log x \log(\log x) \in \mathcal{S} \setminus \mathcal{S}^*$$

($a \notin \mathcal{S}^*$ by [**11**, Theorem 3.6], since a satisfies the hypothesis of the slow growth rate).
A similar argument to that in Theorem 1.1 extends Theorem 1.5 to multiple averages.

THEOREM 1.6. *Let $d \in \mathbb{N}$, (X_i, T_i) be uniquely ergodic systems with unique T_i-invariant measures μ_i, $a_i \in \mathcal{S}$, $1 \le i \le d$ with $a_d \prec \cdots \prec a_1$ and $a_d' \prec \cdots \prec a_1'$, and let f_i be continuous functions on X_i, $1 \le i \le d$. Then, for any coupling ν of these systems, the averages*

$$\frac{1}{N} \sum_{n=0}^{N-1} T_1^{[a_1(n)]} f_1(x_1) \cdots T_d^{[a_d(n)]} f_d(x_d)$$

converge as $N \to \infty$ for ν-a.e. $(x_1, \ldots, x_d) \in X_1 \times \cdots \times X_d$ to

$$\int f_1 \, d\mu_1 \cdots \int f_d \, d\mu_d.$$

Remark. For pointwise averages with a_1 being a linear function, a similar statement to Theorem 1.2 can be derived by assuming unique ergodicity of (X_i, μ_i, T_i) for $i \ge 2$, taking averages over functions in $\mathcal{S}$ on continuous functions (the unique ergodicity of (X_1, μ_1, T_1) is not necessary since, for the linear a_1, we always have single convergence).

In §2, we introduce a specific large family $\mathcal{T}$ of sublinear functions which contains properly the Hardy field functions a with $x^\varepsilon \prec a(x) \prec x$ for some $\varepsilon > 0$ and is contained properly in the class of Fejér functions (see §2 or the Appendix for definitions of these two important classes of functions). We also show (in §§2 and 4) that $\mathcal{T}$ is properly contained in $\mathcal{S}^*$ and that, for functions of different growth rate in $\mathcal{T}$, we get a different growth rate for their derivatives; as a result, our results hold for $a_i \in \mathcal{T}$ of different growth (see Corollary 4.2 for details).

1.3. *Definitions and notation.* A *measure-preserving system* $(X, \mathcal{B}, \mu, T_1, \ldots, T_d)$ is a probability space $(X, \mathcal{B}, \mu)$ endowed with measure-preserving transformations $T_i : X \to X$, meaning that $\mu(T_i^{-1} A) = \mu(A)$ for all $A \in \mathcal{B}$ and $1 \le i \le d$. We omit writing the

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

σ-algebra $\mathcal{X}$ when this causes no confusion. Throughout the paper, we always assume that X is a compact metric space, $\mathcal{B}$ is its Borel σ-algebra and μ is a Borel measure. We let $M(X)$ denote the convex set of probability measures on X which is compact for the weak*-topology. A *coupling* λ of two probability spaces (X_1, μ_1) and (X_2, μ_2) is a measure in $M(X \times Y)$, whose marginals are equal to μ_1 and μ_2, respectively. A *joining* of two measure-preserving systems $(X_1, \mu_1, T_1, \ldots, T_d)$ and $(X_2, \mu_2, S_1, \ldots, S_d)$ is a coupling of (X_1, μ_1) and (X_2, μ_2) that is invariant under the diagonal transformations $T_1 \times S_1, \ldots, T_d \times S_d$ (these definitions extend naturally to k systems, $k \geq 2$). A *factor map* between two measure-preserving systems $(X, \mu, T_1, \ldots, T_d)$ and $(Y, \nu, S_1, \ldots, S_d)$ is a measurable function $\pi : X \to Y$ such that the push-forward measure $\pi_* \mu$ is equal to ν and $\pi \circ T_i = S_i \circ \pi$, $1 \leq i \leq d$. Finally, for two quantities a and b, we write $a \ll b$ if there exists $c > 0$ such that $|a| \leq c \cdot |b|$, and $a \ll_{\delta_1,\ldots,\delta_r} b$ if there exists $c \equiv c(\delta_1, \ldots, \delta_r) > 0$ with $|a| \leq c \cdot |b|$.

In this paper, we study various classes of functions, whose definitions spread throughout the article. For the reader's convenience, we summarize all these definitions, and their connections, in the Appendix.

2. *A nice class of sublinear functions*

In this section, we define a nice class of sublinear functions, $\mathcal{T}$, that first appeared in [**9**], which we will show is a subclass of $\mathcal{S}^*$ defined in the previous section. More specifically, we show, in this section, that $\mathcal{T}$ is a proper subset of $\mathcal{S}$, and then, in §4, we prove that $\mathcal{T}$ is a proper subset of $\mathcal{S}^*$. Let

$$\mathcal{R} := \left\{ a \in \mathcal{C}^3(\mathbb{R}^+) : \text{the limits } \lim_{x\to\infty} \frac{xa'(x)}{a(x)}, \lim_{x\to\infty} \frac{xa''(x)}{a'(x)}, \right. \\ \left. \text{and } \lim_{x\to\infty} \frac{xa'''(x)}{a''(x)} \text{ exist in } \mathbb{R} \right\}$$

and

$$\mathcal{T} := \left\{ a \in \mathcal{R} : \lim_{x\to\infty} \frac{xa'(x)}{a(x)} \in (0, 1) \text{ or } \lim_{x\to\infty} \frac{xa'(x)}{a(x)} = 1 \right. \\ \left. \text{and } \lim_{x\to\infty} a'(x) = 0 \text{ monotonically} \right\}.$$

We start with the connection between $\mathcal{T}$ and some important classes of sublinear functions.

2.0.1. *Fejér functions.* A function $a \in C^1((c, \infty))$, $c \geq 0$, is a *Fejér* function if: (i) $a'(x)$ tends monotonically to 0 as $x \to \infty$; and (ii) $\lim_{x\to\infty} x|a'(x)| = \infty$. We denote with $\mathcal{F}$ the set of all Fejér functions. Note that every $a \in \mathcal{F}$ is eventually monotone and satisfies the growth rate conditions $\log x \prec a(x) \prec x$ (see [**9**] for more details on Fejér functions).

Remark. Since we can find Fejér functions that are not even $\mathcal{C}^2$, we have $\mathcal{S} \subsetneq \mathcal{F}$.

2.0.2. *Hardy field functions.* Let B be the collection of equivalence classes of real-valued functions defined on some halfline (c, ∞), $c \geq 0$, where two functions that agree

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

eventually are identified. These equivalence classes are called *germs* of functions. A *Hardy field* is a subfield of the ring $(B, +, \cdot)$ that is closed under differentiation†. (See [**21**] for more details on Hardy field functions.)

If $\mathcal{H}$ is the union of all Hardy fields, every element of $\mathcal{H}$ has eventually constant sign, from which it follows that if $a \in \mathcal{H}$, then a is eventually monotone and the limit $\lim_{x\to\infty} a(x)$ exists (possibly infinite, as in the $\mathcal{LE}_\varepsilon$ class of functions). For functions $a, b \in \mathcal{H}$ with $b \neq 0$, it follows that the asymptotic growth ratio $\lim_{x\to\infty} a(x)/b(x)$ exists (possibly infinite), a fact that will often justify the use of L'Hospital's rule. For some $\varepsilon > 0$, we denote with $\mathcal{H}_\varepsilon$ the set of functions a which belong to some Hardy field $\mathcal{H}$ and satisfy $x^\varepsilon \prec a(x) \prec x$‡.

Remark. $\mathcal{H}_\varepsilon$ is a proper subset of $\mathcal{T}$ and $\mathcal{T}$ is a proper subset of $\mathcal{F}$.

Indeed, if $a \in \mathcal{H}_\varepsilon$, then $a'(x) \to 0$ monotonically and

$$\lim_{x\to\infty} \frac{xa'(x)}{a(x)} = \lim_{x\to\infty} \frac{\log |a(x)|}{\log x} \in [\varepsilon, 1],$$

so $a \in \mathcal{T}$ (the other limits exist by the properties of a Hardy field). If $a \in \mathcal{T}$ and $\lim_{x\to\infty} (xa'(x)/a(x)) \in (0, 1)$, then $a \in \mathcal{F}$ from [**9**, Lemma 2.1], while if $\lim_{x\to\infty} (xa'(x)/a(x)) = 1$ and $a'(x) \to 0$ monotonically, then $a \in \mathcal{F}$ by [**9**, Lemma 2.2].

Since $\log^\alpha x \in \mathcal{F}\backslash\mathcal{T}$ for all $\alpha > 1$ and $a(x) = x^{1/2}(2 + \cos\sqrt{\log x}) \in \mathcal{T}\backslash\mathcal{H}_\varepsilon$ for all $\varepsilon > 0$ (the derivative of a^2 is not (eventually) monotone by [**9**, §1]), the claim follows.

The following lemma provides a sufficient condition for a function to belong to D_0 (recall the definition from §1).

LEMMA 2.1. *Let $a \in C^1(\mathbb{R})$ and $\alpha, \beta \in \mathbb{R}$ with $\alpha \leq xa'(x)/a(x) \leq \beta$ eventually. Then for any $H > 0$ and large enough x,*

$$\sup_{h\in[-H,H]} \left|\frac{a'(x+h)}{a(x)}\right| \ll_{\alpha,\beta} |x|^{\beta-\alpha-1}. \tag{5}$$

Proof. Since $\alpha \leq xa'(x)/a(x) \leq \beta$ eventually, following the proof of [**9**, Lemma 2.2], we have that there exist positive constants C_α and C_β such that $C_\alpha|x|^\alpha \leq |a(x)| \leq C_\beta|x|^\beta$. Using the fact that eventually we have $|a'(x)/a(x)| \leq \max\{|\alpha|, |\beta|\}/|x|$, we get

$$\sup_{h\in[-H,H]} \left|\frac{a'(x+h)}{a(x)}\right| \leq \max\{|\alpha|, |\beta|\} C_\alpha^{-1} C_\beta \cdot |x|^{\beta-\alpha-1} \times \max\{|1 - H/x|^{\beta-1}, |1 + H/x|^{\beta-1}\}$$

and the result follows. □

Remark. If the constants α, β of Lemma 2.1 satisfy $\beta - \alpha - 1 < 0$ (a special case of this is when $\lim_{x\to\pm\infty} (xa'(x)/a(x)) \in \mathbb{R}$), then the limit of (5) exists (as $x \to \infty$ or $-\infty$) and it is zero.

† We use the word *function* when we refer to elements of B (understanding that all the operations defined and statements made for elements of B are considered only for sufficiently large values of $x \in \mathbb{R}^+$).

‡ We say that the Hardy field functions a that satisfy $\log x \prec a(x) \prec x$ are of *polynomial degree zero*.

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

Since $\mathcal{T} \subseteq \mathcal{F}$, every $a \in \mathcal{T}$ is sublinear, (eventually) monotone and has the property that $1/a'(x) \prec x$. Moreover, we have the following proposition.

PROPOSITION 2.2. *$\mathcal{T}$ is a proper subset of $\mathcal{S}$.*

Proof. Assuming that $a \in \mathcal{T}$ is eventually positive (the other case follows analogously), and using the fact that $\lim_{x\to\infty}(xa'(x)/a(x)) = \alpha \in (0, 1]$, we get $\lim_{x\to\infty}(xa''(x)/a'(x)) = \alpha - 1$ and $\lim_{x\to\infty}(xa'''(x)/a''(x)) = \alpha - 2$ (the properties of $\mathcal{T}$ allow us to use L'Hospital's rule), so, we have that

$$\lim_{x\to\infty}\frac{x(a^{-1})'(x)}{a^{-1}(x)} = \lim_{y\to\infty}\frac{a(y)}{ya'(y)} = \frac{1}{\alpha},$$

$$\lim_{x\to\infty}\frac{x(a^{-1})''(x)}{(a^{-1})'(x)} = -\lim_{y\to\infty}\frac{a''(y)a(y)}{(a'(y))^2} = \frac{1}{\alpha} - 1,$$

and

$$\lim_{x\to\infty}\frac{x(a^{-1})'''(x)}{(a^{-1})''(x)} = \lim_{y\to\infty}\left(\frac{a(y)a'''(y)}{a'(y)a''(y)} - 3\frac{a(y)a''(y)}{(a'(y))^2}\right) = \frac{1}{\alpha} - 2.$$

The previous remark implies that $a \in \mathcal{S}$. Since $a(x) = \log x \log(\log x) \in \mathcal{S}\backslash\mathcal{T}$ (a satisfies all the properties of S—we skip all the elementary calculations—and $xa'(x)/a(x) \to 0$ as $x \to \infty$), we have the claim. □

Remark. The class $\mathcal{T}$ misses not only slow functions as $a_1(x) = \log x \log(\log x)$ from S (i.e., functions a with $1 \prec a \prec x^{\varepsilon}$ for all $\varepsilon > 0$) but also functions as

$$a_2(x) = x^{\alpha}(4/\alpha + \sin\log x)^3$$

for $0 < \alpha < 1/20$ for which we have $a_2 \notin \mathcal{T}$, since the ratio $xa_2'(x)/a_2(x)$ does not have a limit as $x \to \infty$ (it is bounded, though, between positive numbers since $\alpha - 3\alpha/(4-\alpha) \leq xa_2'(x)/a_2(x) \leq \alpha + 3\alpha/(4-\alpha)$), and a_2 is not slow since $x^{\alpha/2} \prec a_2(x)$. Since $a_2(x) \prec x$, $a_2 \in M_1 \subseteq M_0$ with $a_2'(x) \to 0$ and $1/a_2'(x) \prec x$, we actually have that $a_2 \in \mathcal{F}\backslash\mathcal{T}$. We also have that $a_2^{-1} \in D_0 \cap D_1 \cap D_2$ (the lim sup that appear in these sets are limits and equal to zero—we skip all the elementary calculations), and hence $a_2(x) \in \mathcal{S}\backslash\mathcal{T}$.

We will actually show, in §4, that a_2 is a special function as $a_2 \in S^*$.

The following lemma informs us that, for a function in D_0, the ratios of horizontal translations are bounded.

LEMMA 2.3. *Let $a \in \mathcal{C}^1(\mathbb{R})$ and $H > 0$ with* $\limsup_{x\to\infty}\sup_{h\in[-H,H]}|a'(x+h)/a(x)| < \infty$ *(respectively, $x \to -\infty$). Then the quantities $|a(x+\rho)/a(x)|$ are eventually bounded for all $-H \leq \rho \leq H$.*

Proof. Let $\rho \in [-H, H]$. The mean value theorem furnishes a point $h_\rho \in [-H, H]$ with

$$\left|\frac{a(x+\rho)}{a(x)} - 1\right| = \left|\frac{a'(x+h_\rho)}{a(x)}\right|,$$

from which the result follows by our hypothesis. □

We close this section with a fact about the sets D_k. We show that, for a function of S, if the lim sup that appears in D_0 is a limit, then it has to be equal to zero.

PROPOSITION 2.4. *If $a \in S$ with $\lim_{x\to\infty} \sup_{h\in[-1,1]} |(a^{-1})'(x+h)/a^{-1}(x)| = \alpha \in \mathbb{R}$, then $\alpha = 0$.*

Proof. We assume, to the contrary, that $\alpha > 0$ (the case where $a < 0$ is analogous) and we let $0 < \varepsilon < \alpha$. Using the hypothesis, there exists $M > 0$ such that $(a^{-1})'(x+1)/a^{-1}(x) > \alpha - \varepsilon$ for all $x > M$ or, equivalently,

$$\frac{1}{a^{-1}(x+1)a'(a^{-1}(x+1))} > (\alpha - \varepsilon)\frac{a^{-1}(x)}{a^{-1}(x+1)},$$

from which (using the fact that $1/a'(x) \prec x$ and by taking the lim sup) we have that $\lim_{x\to\infty} ((a^{-1}(x-1))/a^{-1}(x)) = 0$. By the mean value theorem, there exists $\xi_x \in [x-1, x]$ with

$$\left|\frac{a^{-1}(x-1)}{a^{-1}(x)} - 1\right| = \frac{(a^{-1})'(\xi_x)}{a^{-1}(x)} \le \frac{(a^{-1})'(x+1)}{a^{-1}(x)}.$$

Letting $x \to \infty$, we get $\alpha \ge 1$. By Lemma 2.3, there exists $c > 0$ such that

$$\frac{(a^{-1})'(x)}{a^{-1}(x)} \ge c\frac{(a^{-1})'(x+1)}{a^{-1}(x)} > c(\alpha - \varepsilon).$$

By integrating and solving this relation for $a(x)$, we get $a(x) \le c_1 \log x + c_2$ for some c_1, c_2 constants with $c_1 > 0$. Then

$$0 < \frac{1}{c_1} \le \lim_{x\to\infty} \frac{\log x}{a(x)} = \lim_{x\to\infty} \frac{1}{xa'(x)} = 0,$$

which is a contradiction. The claim now follows. □

3. *The key invariant properties*

In the following two subsections, we state and prove the invariant arguments that play central roles in our study and will be used in the proof of our main results.

3.1. *The sublinear case.*

In this subsection, we develop the main tool, Lemma 3.1, in order to prove Theorem 1.4, which we will use in the proof of Theorem 1.1.

Recall, from §1, that $\lambda_{\vec{x}}$ is any weak limit in $M(X_1 \times \cdots \times X_d)$ of

$$\lambda_{N,\vec{x}} = \frac{1}{N}\sum_{n=0}^{N-1} (T_1^{[a_1(n)]} \times \cdots \times T_d^{[a_d(n)]})\delta_{\vec{x}}.$$

LEMMA 3.1. *Let $d \in \mathbb{N}$ and $a_i \in \mathcal{S}$, $1 \le i \le d$, with $a_d \prec \cdots \prec a_1$ and $a_d' \prec \cdots \prec a_1'$. Then $\lambda_{\vec{x}}$ is invariant under $T_1 \times \mathrm{id} \times \cdots \times \mathrm{id}$.*

Proof. We can assume without loss of generality that a_1 is eventually positive (the other case is analogous). For $\vec{b} = (b_1, b_2, \ldots, b_d) \in \mathbb{Z}^d$, write $\vec{b}_* = (b_2, \ldots, b_d)$ and let

$$\mathcal{U}_{b_1,\vec{b}_*} := |\{n \in \{1, \ldots, N\} : b_i \le a_i(n) < b_i + 1 \ \forall 1 \le i \le d\}|,$$
$$\delta_{b_1,\vec{b}_*} := \delta_{T_1^{b_1}x_1} \times \cdots \times \delta_{T_d^{b_d}x_d}.$$

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

With this notation, we have that $\lambda_{N,\vec{x}} = (1/N)\sum_{(b_1,\vec{b}_*):\mathcal{U}_{b_1,\vec{b}_*}\neq 0}\mathcal{U}_{b_1,\vec{b}_*}\delta_{b_1,\vec{b}_*}$, and if

$$\tilde{\lambda}_{N,\vec{x}} := (T_1 \times \mathrm{id} \times \cdots \times \mathrm{id})(\lambda_{N,\vec{x}}) = \frac{1}{N}\sum_{(b_1,\vec{b}_*):\mathcal{U}_{b_1-1,\vec{b}_*}\neq 0}\mathcal{U}_{b_1-1,\vec{b}_*}\delta_{b_1,\vec{b}_*},$$

then

$$\lambda_{N,\vec{x}} - \tilde{\lambda}_{N,\vec{x}} = \frac{1}{N}\sum_{(b_1,\vec{b}_*):\mathcal{U}_{b_1,\vec{b}_*}\neq 0,\mathcal{U}_{b_1-1,\vec{b}_*}\neq 0}(\mathcal{U}_{b_1,\vec{b}_*} - \mathcal{U}_{b_1-1,\vec{b}_*})\delta_{b_1,\vec{b}_*} \tag{6}$$

$$+ \frac{1}{N}\sum_{(b_1,\vec{b}_*):\mathcal{U}_{b_1,\vec{b}_*}\neq 0,\mathcal{U}_{b_1-1,\vec{b}_*}=0}\mathcal{U}_{b_1,\vec{b}_*}\delta_{b_1,\vec{b}_*} \tag{7}$$

$$- \frac{1}{N}\sum_{(b_1,\vec{b}_*):\mathcal{U}_{b_1,\vec{b}_*}=0,\mathcal{U}_{b_1-1,\vec{b}_*}\neq 0}\mathcal{U}_{b_1-1,\vec{b}_*}\delta_{b_1,\vec{b}_*}. \tag{8}$$

We will study each term separately. Terms (7) and (8) are treated similarly, and so we only study the terms (6) and (7).

Term (6): For $1 \le i \le d$, let

$$C_{a_i,b_i} := \min\{a_i^{-1}(b_i), a_i^{-1}(b_i+1)\} \quad \text{and} \quad C'_{a_i,b_i} := \max\{a_i^{-1}(b_i), a_i^{-1}(b_i+1)\}.$$

The conditions $\mathcal{U}_{b_1,\vec{b}_*} \neq 0$ and $\mathcal{U}_{b_1-1,\vec{b}_*} \neq 0$ imply that

$$\max_{1\le i\le d}\{C_{a_i,b_i}\} < \min_{1\le i\le d}\{C'_{a_i,b_i}\} \quad \text{and} \tag{9}$$

$$\max_{2\le i\le d}\{C_{a_1,b_1-1}, C_{a_i,b_i}\} < \min_{2\le i\le d}\{C'_{a_1,b_1-1}, C'_{a_i,b_i}\}. \tag{10}$$

Using relations (9) and (10), we get

$$C_{a_i,b_i} < a_1^{-1}(b_1) < C'_{a_i,b_i} \quad \text{for all } 2 \le i \le d,$$

from which we have

$$b_i < a_i \circ a_1^{-1}(b_1) < b_i + 1,$$

and hence $b_i = [a_i \circ a_1^{-1}(b_1)]$ for all $2 \le i \le d$.

By the mean value theorem, for some $\xi_b \in (b_1 - 1, b_1 + 1)$,

$$|\mathcal{U}_{b_1,\vec{b}_*} - \mathcal{U}_{b_1-1,\vec{b}_*}| \le |a_1^{-1}(b_1+1) - 2a_1^{-1}(b_1) + a_1^{-1}(b_1-1)| + 2 = |(a_1^{-1})''(\xi_b)| + 2.$$

Since $a_1^{-1} \in D_2 \cup M_2$ for large enough b_1, we have that $|(a_1^{-1})''(\xi_b)| \ll |(a_1^{-1})''(b_1)|$ (or $|(a_1^{-1})''(b_1 \pm 1)|$, which is studied similarly). So,

$$\frac{1}{N}\sum_{b_1=1}^{[a_1(N)]-1}|\mathcal{U}_{b_1,\vec{b}_*} - \mathcal{U}_{b_1-1,\vec{b}_*}| \ll \frac{1}{N}\sum_{b_1=1}^{[a_1(N)]-1}|(a_1^{-1})''(b_1)|$$

(the term $2[a_1(N)]/N$ is null by sublinearity of a_1). By the same property, for large enough b_1 and all $t \in [b_1, b_1+1]$, we have that $|(a_1^{-1})''(b_1)| \ll |(a_1^{-1})''(t)|$, so

$$|(a_1^{-1})''(b_1)| \ll \int_{b_1}^{b_1+1}|(a_1^{-1})''(t)|\,dt.$$

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

Hence,

$$\frac{1}{N}\sum_{b_1=1}^{[a_1(N)]-1}|(a_1^{-1})''(b_1)| \ll \frac{1}{N}\int_*^{[a_1(N)]}|(a_1^{-1})''(t)|\,dt \ll \frac{|(a_1^{-1})'([a_1(N)])|}{N} \tag{11}$$

(where we used the fact that $a_1^{-1}\in D_2$ so $(a_1^{-1})''$ has constant sign—in our case, here, is positive). Using the fact that $a_1^{-1}\in D_1$, we have that $(a_1^{-1})'([a_1(N)])/(a_1^{-1})'(a_1(N))$ is bounded for large N, so the right-hand side of (11) is bounded by a constant multiple of $(Na_1'(N))^{-1}$, which goes to zero as $N\to\infty$ and, consequently, term (6) goes to zero as $N\to\infty$.

Term (7): The conditions $\mathcal{U}_{b_1,\vec{b}_*}\neq 0$ and $\mathcal{U}_{b_1-1,\vec{b}_*}=0$ imply that

$$\max_{1\le i\le d}\{C_{a_i,b_i}\} < \min_{1\le i\le d}\{C'_{a_i,b_i}\} \quad \text{and} \tag{12}$$

$$\text{as in (10) but with } C_{a_i,b_i} > a_1^{-1}(b_1)-1 \quad \text{for at least one } 2\le i\le d. \tag{13}$$

If $\{i_0=1, i_1, \ldots, i_r\}$ is the set of indices for which $C_{a_i,b_i} > a_1^{-1}(b_1)-1$, let

$$B_{\{i_0,i_1,\ldots,i_r\}} := \{\vec{b}_* : a_1^{-1}(b_1)\le C_{a_{i_1},b_{i_1}} \le \cdots \le C_{a_{i_r},b_{i_r}}\}.$$

Then

$$a_1^{-1}(b_1)-1 < C_{a_{i_j},b_{i_j}} \le C_{a_{i_r},b_{i_r}} < C'_{a_{i_j},b_{i_j}} \quad \text{for all } 0<j\le r,$$

from which we get

$$b_{i_j} \le a_{i_j}\circ a_{i_r}^{-1}(b_{i_r}+e_{i_r}) < b_{i_j}+1$$

for some $e_{i_r}\in\{0,1\}$. Hence, $b_{i_j}=[a_{i_j}\circ a_{i_r}^{-1}(b_{i_r}+e_{i_r})]$ for all $0<j\le r$.

For $j=0$,

$$a_1^{-1}(b_1)-1 < C_{a_{i_r},b_{i_r}} < a_1^{-1}(b_1+1)$$

or

$$a_1(C_{a_{i_r},b_{i_r}})-1 < b_1 < a_1(C_{a_{i_r},b_{i_r}}+1).$$

So, for each b_{i_r} fixed, $b_1=[a_1(C_{a_{i_r},b_{i_r}})]$ or $b_1=[a_1(C_{a_{i_r},b_{i_r}}+1)]$.

For $i\notin\{i_1,\ldots,i_r\}$,

$$C_{a_i,b_i}\le a_1^{-1}(b_1) < C_{a_{i_r},b_{i_r}} < C'_{a_i,b_i}$$

or

$$b_i \le a_i\circ a_{i_r}^{-1}(b_{i_r}+e_{i_r}) < b_i+1$$

for some $e_{i_r}\in\{0,1\}$, and hence $b_i=[a_i\circ a_{i_r}^{-1}(b_{i_r}+e_{i_r})]$.

The average of $|\mathcal{U}_{b_1,\vec{b}_*}|$ can be split into finitely many sums over $(b_1,\vec{b}_*)$ for $\vec{b}_*\in B_{\{i_0,i_1,\ldots,i_r\}}$ for some $\{i_0,i_1,\ldots,i_r\}$. For $\vec{b}_*\in B_{\{i_0,i_1,\ldots,i_r\}}$, $|\mathcal{U}_{b_1,\vec{b}_*}|$ is bounded by

$$|a_1^{-1}(b_1+1)-a_1^{-1}(b_1)|+1 \ll |(a_1^{-1})'(b_1)|+1 = |(a_1^{-1})'([a_1\circ a_{i_r}^{-1}(b_{i_r}+e_{i_r})])|+1.$$

This approximation follows by the mean value theorem and the fact that $a_1^{-1}\in D_1$. Hence, every average is estimated by a constant multiple of

$$\frac{1}{N}\left|\int_*^{[a_{i_r}(N)]}|(a_1^{-1})'([a_1\circ a_{i_r}^{-1}(t)])|\,dt\right| \ll \frac{1}{N}\left|\int_*^{[a_{i_r}(N)]}|(a_1^{-1})'(a_1\circ a_{i_r}^{-1}(t))|\,dt\right|, \tag{14}$$

where we used the fact that $a_{i_r} \prec a_1$ and the sublinearity of a_{i_r}. So, the right-hand side of (14) is bounded by

$$\frac{1}{N}\left|\int_*^{[a_{i_r}(N)]} \frac{1}{|a_1'(a_{i_r}^{-1}(t))|}\,dt\right| = \frac{1}{N}\left|\int_*^{a_{i_r}^{-1}([a_{i_r}(N)])} \left|\frac{a_{i_r}'(t)}{a_1'(t)}\right|\,dt\right|.$$

Since $a_{i_r}^{-1} \in D_0$, we have $a_{i_r}^{-1}([a_{i_r}(N)])/N \to 1$ as $N \to \infty$, and hence the last integral goes to zero, since the integrand is arbitrarily small for large N because of $a_{i_r}' \prec a_1'$.

This proves that terms (7) and (8) go to zero as $N \to \infty$ and we get the conclusion. □

3.2. *Linearity of the fastest term.* In this subsection, we deal with the case where the fastest function is linear. If the leading coefficient is rational, we can work as in Lemma 3.1, while if the leading coefficient is irrational, with an explicit example (see the remark after Lemma 3.2), we show that the invariant property of Lemma 3.1 can fail, but one can still deal with this case via the suspension flow and Birkhoff's ergodic theorem (see §4.1).

3.2.1. *Rational leading coefficient.* If $a_1(n) = kn + \ell$, with $k \in \mathbb{Q}\backslash\{0\}$, $\ell \in \mathbb{R}$, it suffices to cover the case where $a_1(n) = pn$ with $p \in \mathbb{Z}\backslash\{0\}$ (see details in the proof of Theorem 1.2). As before, let $\lambda_{\vec{x}}$ be a weak limit of $\lambda_{N,\vec{x}}$.

LEMMA 3.2. *For $d \in \mathbb{N}$, let $a_1(n) = pn$, $p \in \mathbb{Z}\backslash\{0\}$ and a_i, $2 \le i \le d$, be functions with $a_d \prec \cdots \prec a_1$. Then $\lambda_{\vec{x}}$ is invariant under $T_1^p \times \mathrm{id} \times \cdots \times \mathrm{id}$.*

Proof. We can assume that $p > 0$ (the case where $p < 0$ is analogous), and hence a_1 is eventually positive. For $\vec{b} = (b_1, b_2, \ldots, b_d) \in \mathbb{Z}^d$ we write $\vec{b}_* = (b_2, \ldots, b_d)$ and we set

$$\mathcal{U}_{b_1,\vec{b}_*} = \begin{cases} 1 & \text{if } [a_i(b_1)] = b_i \quad \text{for all } 2 \le i \le d, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{and}$$

$$\delta_{b_1,\vec{b}_*} = \delta_{T_1^{[a_1(b_1)]}x_1} \times \delta_{T_2^{b_2}x_2} \times \cdots \times \delta_{T_d^{b_d}x_d}$$

(note that the position of b_1 here is not the same as in the proof of Lemma 3.1).

With this notation, we have that $\lambda_{N,\vec{x}} = (1/N)\sum_{(b_1,\vec{b}_*):\mathcal{U}_{b_1,\vec{b}_*}\neq 0} \mathcal{U}_{b_1,\vec{b}_*}\delta_{b_1,\vec{b}_*}$, and

$$\tilde{\lambda}_{N,\vec{x}} := (T_1^p \times \mathrm{id} \times \cdots \times \mathrm{id})(\lambda_{N,x}) = \frac{1}{N} \sum_{(b_1,\vec{b}_*):\mathcal{U}_{b_1-1,\vec{b}_*}\neq 0} \mathcal{U}_{b_1-1,\vec{b}_*}\delta_{b_1,\vec{b}_*},$$

where we used the fact that $[a_1(b_1 - 1)] + p = [a_1(b_1)]$. Then

$$\lambda_{N,\vec{x}} - \tilde{\lambda}_{N,\vec{x}} = \frac{1}{N} \sum_{(b_1,\vec{b}_*):\mathcal{U}_{b_1,\vec{b}_*}\neq 0,\mathcal{U}_{b_1-1,\vec{b}_*}\neq 0} (\mathcal{U}_{b_1,\vec{b}_*} - \mathcal{U}_{b_1-1,\vec{b}_*})\delta_{b_1,\vec{b}_*} \tag{15}$$

$$+ \frac{1}{N} \sum_{(b_1,\vec{b}_*):\mathcal{U}_{b_1,\vec{b}_*}\neq 0,\mathcal{U}_{b_1-1,\vec{b}_*}=0} \mathcal{U}_{b_1,\vec{b}_*}\delta_{b_1,\vec{b}_*} \tag{16}$$

$$- \frac{1}{N} \sum_{(b_1,\vec{b}_*):\mathcal{U}_{b_1,\vec{b}_*}=0,\mathcal{U}_{b_1-1,\vec{b}_*}\neq 0} \mathcal{U}_{b_1-1,\vec{b}_*}\delta_{b_1,\vec{b}_*}. \tag{17}$$

We study the terms (15) and (16) as in the sublinear case.

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

Term (15): Since each $\mathcal{U}_{b_1,b_*}$ is at most one, $\mathcal{U}_{b_1,b_*} = \mathcal{U}_{b_1-1,b_*} = 1$, so term (15) is equal to zero.

Term (16): Let

$$C_{a_i,b_i} := \min\{a_i^{-1}(b_i), a_i^{-1}(b_i+1)\} \quad \text{and} \quad C'_{a_i,b_i} := \max\{a_i^{-1}(b_i), a_i^{-1}(b_i+1)\}.$$

The conditions $\mathcal{U}_{b_1,\vec{b}_*} \neq 0$ and $\mathcal{U}_{b_1-1,\vec{b}_*} = 0$ imply that

$$\max_{1\le i\le d}\{C_{a_i,b_i}\} \le b_1 < \min_{1\le i\le d}\{C'_{a_i,b_i}\} \quad \text{and} \tag{18}$$

$$C_{a_i,b_i} \ge b_1 - 1 \quad \text{for at least one } 2 \le i \le d. \tag{19}$$

If $\{i_1, \ldots, i_r\}$ is the set of indices for which $C_{a_i,b_i} \ge b_1 - 1$, let

$$B_{\{i_1,\ldots,i_r\}} := \{\vec{b}_* : b_1 - 1 \le C_{a_{i_1},b_{i_1}} \le \cdots \le C_{a_{i_r},b_{i_r}}\}.$$

As in Lemma 3.1 we get $b_1 = [C_{a_{i_r},b_{i_r}}]$ or $b_1 = [C_{a_{i_r},b_{i_r}}] + 1$ and $b_i = [a_i \circ a_{i_r}^{-1}(b_{i_r} + e_{i_r})]$, $2 \le i \le d$, for some $e_{i_r} \in \{0, 1\}$.

For fixed b_{i_r}, there are at most 2^r terms $|\mathcal{U}_{b_1,\vec{b}_*}|$ in (16) whose last index is equal to b_{i_r}. Since $|b_{i_r}| \le a_{i_r}(N)$, (16) is bounded by a constant multiple of $a_{i_r}(N)/N$, which goes to zero as $N \to \infty$ and hence the conclusion of the lemma follows. □

Remark. In the previous proof, we essentially used the relation $[a_1(b_1 - 1)] + p = [a_1(b_1)]$, which is not, in general, true for expressions of the form $a(t) = \gamma n + \ell$, $\gamma \notin \mathbb{Q}$. Actually, Lemma 3.2 cannot even be extended to the $d = 1$ case if $a_1(n) = \gamma n + \ell$, $\gamma \notin \mathbb{Q}$.

Indeed, let $d = 1$ and $(\mathbb{T}, \mu)$ be the one-dimensional torus endowed with the Lebesgue measure. Let $T : \mathbb{T} \to \mathbb{T}$ with $Tx = \gamma^{-1} + x$. Then ($\{\cdot\}$ denotes the fractional part)

$$T^{[\gamma n+\ell]}x = \gamma^{-1}[\gamma n + \ell] + x = \gamma^{-1}\ell + x - \gamma^{-1}\{\gamma n\}.$$

Since $(\{\gamma n\})_n$ is equidistributed on $\mathbb{T}$, for all $f \in L^\infty(\mu)$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T^{[\gamma n+\ell]} f(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(\gamma^{-1}\ell + x - \gamma^{-1}\{\gamma n\})$$
$$= \int_{\gamma^{-1}(\ell-1)+x}^{\gamma^{-1}\ell+x} f(y)\, d\mu(y)$$

and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T^{[\gamma n+\ell]+1} f(x) = \int_{\gamma^{-1}+x}^{\gamma^{-1}(\ell+1)+x} f(y)\, d\mu(y).$$

So λ_x is obviously not T-invariant. Nevertheless, we will show in the next section that there is an explicit expression for the limit $\lim_{N\to\infty} (1/N) \sum_{n=0}^{N-1} T^{[\gamma n+\ell]} f(x)$, $\gamma \notin \mathbb{Q}$.

4. *Single pointwise averages*

Let (X, μ, T) be a measure-preserving system. We start this section by justifying why, for $a \in \mathcal{S}^*$, the average $(1/N) \sum_{n=0}^{N-1} T^{[a(n)]} f(x)$ converges to $\mathbb{E}(f \mid \mathcal{I}(T))(x)$ for μ-a.e. $x \in X$ for every bounded measurable function f. Since we assume that the average does converge (by the definition of $\mathcal{S}^*$) it suffices to show that it converges in

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

$L^2(\mu)$ to $\mathbb{E}(f \mid \mathcal{I}(T))$. By writing $f = \mathbb{E}(f \mid \mathcal{I}(T)) + g$, where $g = f - \mathbb{E}(f \mid \mathcal{I}(T))$, it suffices to show that the average goes to zero for a function g with zero conditional expectation with respect to $\mathcal{I}(T)$. To prove that, it suffices to show it for coboundary functions, i.e., for functions of the form $g - g \circ T$. By decomposing $\mu = \int \delta_x \, d\mu(x)$, it is a consequence of Lemma 3.1 that the average does converge to zero in $L^2(\mu)$ for coboundary functions.

We now establish all the pointwise limits of (1) for $d = 1$ for the class $\mathcal{T} \cup \mathcal{L}_*$, where $\mathcal{L}_*$ denotes the set of linear functions with non-zero leading coefficient.

THEOREM 4.1. *Let (X, μ, T) be a measure-preserving system and $f \in L^\infty(\mu)$. If $a \in \mathcal{T}$, then, for μ-a.e. $x \in X$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T^{[a(n)]} f(x) = \mathbb{E}(f \mid \mathcal{I}(T))(x).$$

Theorem 4.1 is proved in [**11**, Theorem 3.4] for $a \in \mathcal{H}_\varepsilon$ but the same result holds for $a \in \mathcal{T}$. For completeness, we sketch the idea of the proof.

By [**9**, Theorem 3.5], if a satisfies some properties, then the average of a sequence $(x_n)_n$ along $([a(n)])_n$, i.e., $(1/N) \sum x_{[a(n)]}$, converges to zero as long as the usual average $(1/N) \sum x_n$ converges to zero. The proof that the assumptions of [**9**, Theorem 3.5] are satisfied for functions from $\mathcal{T}$ can be found in [**9**, Lemma 2.5]. More specifically, [**9**, Lemma 2.5] covers the case where $a > 0$, while the case where $a < 0$ follows by the fact that $[a(n)] = -[-a(n)] - 1$ in a set of density one since $[a(n)] = -[-a(n)]$ only happens when $a(n)$ is an integer, which, up to time N, happens at most $a(N)$ times. The claim now follows by the sublinearity of a.

Remarks.

(1) Combining this result with Proposition 2.2, we get that $\mathcal{T}$ is a subset of $\mathcal{S}^*$. In fact, $\mathcal{S}^*$ properly contains $\mathcal{T}$. Indeed, recall by §2 that, for a small positive α,

$$a_2(x) = x^\alpha (4/\alpha + \sin \log x)^3 \in \mathcal{S} \backslash \mathcal{T}.$$

The function a_2 belongs to $\mathcal{S}^*$ as it satisfies all the conditions of [**9**, Theorem 3.5]. To be more precise, let $\phi(n) = |\{m \in \mathbb{N} : n = [a_2(m)]\}|$ and $\Phi(n) = \sum_{k=0}^{n} \phi(k)$. Then all the following conditions of [**9**, Theorem 3.5] hold:

(i) $\lim_{n\to\infty} \phi(n) = \infty$ (obvious);

(ii) $\phi(n)$ is almost increasing, i.e., $\phi(n) = q_n + p_n$, where $(q_n)_n$ is increasing and $(p_n)_n$ is bounded (this follows by [**9**, Lemma 2.4] since a_2^{-1} is increasing and a_2' is decreasing); and

(iii) $n\phi(n)/\Phi(n) \le c$ for some $c > 0$ (this follows by the proof of [**9**, Lemma 2.5] since $n\phi(n)/\Phi(n)$ is arbitrarily close to $n(a_2^{-1})'(n)/a_2^{-1}(n)$, and hence it is bounded above by a positive number since, as we already mentioned in §2, $na_2'(n)/a_2(n)$ is bounded below by $\alpha - 3\alpha/4 - \alpha > 0$).

Also, it is worth recalling at this point that Theorem 1.5 establishes the same result as Theorem 4.1 for a larger class of functions in the case of uniquely ergodic systems.

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

(2) Summarizing the information that we have for the classes of interest, gives

$$\mathcal{H}_\varepsilon \subsetneq \mathcal{T} \subsetneq \mathcal{S}^* \subsetneq \mathcal{S} \subsetneq \mathcal{F} \subsetneq \mathcal{SL},$$

where the strictness of the inclusions are given by: $x^{1/2}(2 + \cos\sqrt{\log x})$; $x^\alpha(4/\alpha + \sin\log x)^3$ for $0 < \alpha < 1/20$; $\log x \log(\log x)$; any Fejér non-$\mathcal{C}^2$ function; and $\log x$, respectively.

We can now show that Theorem 1.4 holds for the class $\mathcal{T}$.

COROLLARY 4.2. *Let $d \in \mathbb{N}$, (X_i, μ_i, T_i), be measure-preserving systems, $a_i \in \mathcal{T}$, $1 \le i \le d$ with $a_d \prec \cdots \prec a_1$ and let ν be any coupling of the spaces (X_i, μ_i). Then, for ν-a.e. $\vec{x} \in X_1 \times \cdots \times X_d$, we have that $\lambda_{N,\vec{x}}$ converges to $\mu_{[T_1,\ldots,T_d],\vec{x}}$ as $N \to \infty$.*

This corollary follows (together with the analogous results of Theorems 1.1, 1.2, 1.5 and 1.6 with the condition $a_i \in \mathcal{S}^*$, or $\mathcal{S}$, $a_d \prec \cdots \prec a_1$ and $a'_d \prec \cdots \prec a'_1$ replaced by $a_i \in \mathcal{T}$, or $\mathcal{T} \cup \mathcal{L}_*$ in Theorem 1.2, and $a_d \prec \cdots \prec a_1$) by the fact that $\mathcal{T} \subseteq \mathcal{S}^*$ and that if $a_1, a_2 \in \mathcal{T} \cup \mathcal{L}_*$ with $a_2 \prec a_1$ (so $a_2 \in \mathcal{T}$), then we have that $a'_2 \prec a'_1$ by the identity

$$\frac{a'_2(x)}{a'_1(x)} = \frac{xa'_2(x)}{a_2(x)} \cdot \frac{a_1(x)}{xa'_1(x)} \cdot \frac{a_2(x)}{a_1(x)}.$$

Of course, for functions of different growth $a_1 \prec a_2$ in S^* in general, if some of the limits $xa'_i(x)/a_i(x)$ do not exist as $x \to \infty$ (as in the case of the function that we saw in the remarks of this section) we do not expect to get the different growth rate of the derivatives, so we have to assume it.

If, for every $\alpha \in (0, 1]$, we set $\mathcal{T}(\alpha) := \{a \in \mathcal{T} : \lim_{x\to\infty}(xa'(x)/a(x)) = \alpha\}$, then the following remark gives us a relation between the growth rate of $a_i \in \mathcal{T}(\alpha_i)$, $i = 1, 2$, and α_1, α_2.

Remark. If $a_i \in \mathcal{T}(\alpha_i)$, $i = 1, 2$, with $a_2 \prec a_1$, then $\alpha_2 \le \alpha_1$.

Indeed, this follows by [**9**, Lemmas 2.1 and 2.6]. Let $a = a_2/a_1$ and note that $\lim_{x\to\infty}(xa'(x)/a(x)) = \alpha_2 - \alpha_1$. If $\alpha_2 > \alpha_1$, then, by the argument in [**9**, Lemma 2.1], we have that $|a(x)| \to \infty$ as $x \to \infty$, which is a contradiction (note that we only used here the fact that a is bounded).

Conversely, if $\alpha_2 < \alpha_1$, then $a_2 \prec a_1$.

The same argument, for $1/a$, gives us that $a(x) \to 0$ as $x \to \infty$, and hence $a_2 \prec a_1$.

Note, at this point, that it can happen that $a_2 \prec a_1$ while both functions belong to the same $\mathcal{T}(\alpha)$, as $a_2(x) = x^{1/3}$ and $a_1(x) = x^{1/3}\log x$, where $a_2 \prec a_1$ with $a_1, a_2 \in \mathcal{T}(1/3)$.

We now deal with the case of an irrational leading coefficient in the class $\mathcal{L}_*$.

4.1. *Irrational leading coefficient.* In this subsection, we study the limit along linear iterates of the form $\lim_{N\to\infty}(1/N)\sum_{n=0}^{N-1} T^{[\gamma n+\ell]}f(x)$ when γ is irrational. To do this, we have to introduce some additional notions and tools.

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

4.1.1. *A generalized ergodic theorem.* Let (X, μ, T) be a measure-preserving system, $\mathcal{I}_{\gamma,m}(T)$ be the sub-σ-algebra generated by the eigenspace of T with eigenvalue $-m/\gamma$, $\mathcal{U}_{\gamma,m} \subseteq L^2(\mu)$ be the closed subalgebra generated by all functions of the form $Tg - \exp(2\pi i(m/\gamma))g$, $\mathcal{I}_\gamma(T) = \bigvee_{m\in\mathbb{Z}} \mathcal{I}_{\gamma,m}(T)$ and $\mathcal{U}_\gamma(T) = \bigcap_{m\in\mathbb{Z}} \mathcal{U}_{\gamma,m}(T)$. We have the following structure theorem (its proof is routine and we omit it).

THEOREM 4.3. *Let (X, μ, T) be a measure-preserving system and $\gamma \in \mathbb{R}\backslash\mathbb{Q}$. Then*

$$L^2(\mu) = \mathcal{I}_{\gamma,m}(T) \oplus \mathcal{U}_{\gamma,m}(T)$$

for all $m \in \mathbb{Z}$. In particular,

$$L^2(\mu) = \mathcal{I}_\gamma(T) \oplus \mathcal{U}_\gamma(T).$$

We also have the following von Neumann-type mean ergodic theorem.

PROPOSITION 4.4. *Let (X, μ, T) be a measure-preserving system, $\gamma \in \mathbb{R}\backslash\mathbb{Q}$, $\ell \in \mathbb{R}$ and $f \in L^2(\mu)$. We have the following (each convergence takes place in $L^2(\mu)$).*

- *If $\mathbb{E}(f \mid \mathcal{I}_\gamma(T)) = 0$, then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T^{[\gamma n+\ell]} f = 0.$$

- *If f is measurable in $\mathcal{I}_{\gamma,m}(T)$, then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T^{[\gamma n+\ell]} f = \exp\bigg(2\pi i \frac{m\ell}{\gamma}\bigg) \cdot \frac{\exp(-2\pi i(m/\gamma)) - 1}{-2\pi i(m/\gamma)} f.$$

Consequently, for all $f \in L^2(\mu)$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T^{[\gamma n+\ell]} f = \sum_{m\in\mathbb{Z}} \exp\bigg(2\pi i \frac{m\ell}{\gamma}\bigg) \cdot \frac{\exp(-2\pi i(m/\gamma)) - 1}{-2\pi i(m/\gamma)} \mathbb{E}(f \mid I_{\gamma,m}(T)).$$

Proof. Assume that $\mathbb{E}(f \mid \mathcal{I}_\gamma(T)) = 0$. By Theorem 4.3, $f \in \mathcal{U}_\gamma(T)$. Let $\varepsilon > 0$. By definition, for all $j \in \mathbb{Z}$, there exist $g_j, \varepsilon_j \in L^2(\mu)$ such that $f = Tg_j - \exp(2\pi i(j/\gamma))g_j + \varepsilon_j$ and $\|\varepsilon_j\|_2 \leq \varepsilon$. We may assume that $\gamma > 0$ since the proof of the other case is identical. We first assume that $\gamma > 1$. Note that $m = [n\gamma + \ell]$ for some $n \in \mathbb{Z}$ if and only if $\{(m - \ell)/\gamma\} \in (1 - 1/\gamma, 1)$. Additionally, for each $m \in \mathbb{Z}$, there is at most one $n \in \mathbb{Z}$ such that $m = [n\gamma + \ell]$. So,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T^{[\gamma n+\ell]} f = \lim_{N\to\infty} \frac{\gamma}{N} \sum_{n=0}^{N-1} \mathbf{1}_{(1-1/\gamma,1)}\bigg(\bigg\{\frac{n-\ell}{\gamma}\bigg\}\bigg) T^n f.$$

Let

$$\mathbf{1}_{(1-1/\gamma,1)}\bigg(\bigg\{x - \frac{\ell}{\gamma}\bigg\}\bigg) = \sum_{j\in\mathbb{Z}} a_j \exp(-2\pi i j x), \quad a_j \in \mathbb{R}$$

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

be the Fourier expansion of the function $\mathbf{1}_{(1-1/\gamma,1)}(\{\cdot - \ell/\gamma\})$. Then

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} T^{[\gamma n+\ell]} f = \lim_{N\to\infty} \frac{\gamma}{N}\sum_{n=0}^{N-1}\bigg(\sum_{j\in\mathbb{Z}} a_j \exp\bigg(-2\pi i\frac{jn}{\gamma}\bigg)T^n f\bigg)$$

$$= \sum_{j\in\mathbb{Z}} a_j\bigg(\lim_{N\to\infty}\frac{\gamma}{N}\bigg(\sum_{n=1}^{N}\exp\bigg(-2\pi i\frac{j(n-1)}{\gamma}\bigg)T^n g_j$$

$$-\sum_{n=0}^{N-1}\exp\bigg(-2\pi i\frac{j(n-1)}{\gamma}\bigg)T^n g_j\bigg)\bigg)$$

$$+\sum_{j\in\mathbb{Z}} a_j\bigg(\lim_{N\to\infty}\frac{\gamma}{N}\bigg(\sum_{n=1}^{N}\exp\bigg(-2\pi i\frac{j(n-1)}{\gamma}\bigg)T^n \varepsilon_j\bigg)\bigg).$$

The first series is equal to zero, while the second one has L^2-norm smaller than ε. Since ε is arbitrary, we conclude that $\lim_{N\to\infty}(1/N)\sum_{n=0}^{N-1} T^{[\gamma n+\ell]} f = 0$.

If $0<\gamma<1$, pick $W\in\mathbb{N}$ such that $\gamma W>1$. Then, by the previous computation,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} T^{[\gamma n+\ell]} f = \frac{1}{W}\sum_{k=0}^{W-1}\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} T^{[(\gamma W)n+(\gamma k+\ell)]} f = 0.$$

This finishes the proof of the first part. Suppose now that f is measurable in $\mathcal{I}_{\gamma,m}(T)$. Then $Tf(x) = \exp(2\pi i(m/\gamma))f(x)$ for μ-a.e. $x\in X$. For such $x\in X$,

$$\frac{1}{N}\sum_{n=0}^{N-1} T^{[\gamma n+\ell]} f(x) = \frac{1}{N}\sum_{n=0}^{N-1}\exp\bigg(2\pi i\frac{m[n\gamma+\ell]}{\gamma}\bigg)f(x)$$

$$= \exp\bigg(2\pi i\frac{m\ell}{\gamma}\bigg)f(x)\cdot\frac{1}{N}\sum_{n=0}^{N-1}\exp\bigg(-2\pi i\frac{m\{n\gamma+\ell\}}{\gamma}\bigg).$$

Since the sequence $(\{n\gamma+\ell\})_n$ is equidistributed in $\mathbb{T}$,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\exp\bigg(-2\pi i\frac{m\{n\gamma+\ell\}}{\gamma}\bigg) = \int_{\mathbb{T}}\exp\bigg(-2\pi i\frac{mx}{\gamma}\bigg)\,dx$$

$$= \frac{\exp(-2\pi i(m/\gamma))-1}{-2\pi i(m/\gamma)}.$$

This completes the proof. □

4.1.2. *A special extension.* Let (X,μ,T) be a measure-preserving system. We build an extension system of X which will be used to prove Theorem 1.2. Let $([0,1],\mathcal{D},m)$ be the $[0,1]$ interval with the Lebesgue measure m. Let R be the equivalence relation on $[0,1]\times X$ generated by $((1,x),(0,Tx))$, $x\in X$, $Y=([0,1]\times X)/R$, $\mathcal{Y}=(\mathcal{D}\times\mathcal{B})/R$, $\nu = m\times\mu/R$, and $\tilde{T}=\mathrm{id}\times T$. Then we have a factor map $\pi:(Y,\mathcal{Y},\nu,\tilde{T})\to(X,\mathcal{B},\mu,T)$, where π is the projection to the second coordinate. Let $S:Y\to Y$ with $S(t,x)=S(t+\gamma,x)$. Then (using the relations $[\gamma+\{\gamma'\}]+[\gamma']=[\gamma+\gamma']$ and $\{\gamma+\{\gamma'\}\}=\{\gamma+\gamma'\}$),

$$S^n(\{\ell\},T^{[\ell]}x) = (n\gamma+\ell,x) = (\{n\gamma+\ell\},T^{[n\gamma+\ell]}x)$$

for all $n \in \mathbb{Z}$ and $x \in X$. For a function f on X, we define its extension, $\tilde{f}$ on Y, by $\tilde{f}(t, x) = f(x)$ for all $t \in [0, 1)$, $x \in X$. Then

$$T^{[\gamma n+\ell]} f(x) = \tilde{f}(\{n\gamma + \ell\}, T^{[n\gamma+\ell]} x) = S^n \tilde{f}(\{\ell\}, T^{[\ell]} x).$$

COROLLARY 4.5. *Let the quantifiers be as above. Then, for μ-a.e. $x \in X$,*

$$\mathbb{E}(\tilde{f} \mid \mathcal{I}(S))(\{\ell\}, T^{[\ell]} x) = \sum_{m \in \mathbb{Z}} \exp\bigg(2\pi i \frac{m\ell}{\gamma}\bigg) \cdot \frac{\exp(-2\pi i(m/\gamma)) - 1}{-2\pi i(m/\gamma)} \mathbb{E}(f \mid I_{\gamma,m}(T))(x).$$

Proof. By Birkhoff's ergodic theorem, the limit $\lim_{N\to\infty} (1/N) \sum_{n=0}^{N-1} S^n \tilde{f}(y, z)$ exists and is equal to $\mathbb{E}(\tilde{f} \mid \mathcal{I}(S))(y, z)$ for ν-a.e. $(y, z) \in Y$. Since $\tilde{f}(y, z) = \tilde{f}(y', z)$ for all $y, y' \in [0, 1]$, $z \in X$, and μ is T-invariant, it is easy to conclude that $\lim_{N\to\infty} (1/N) \sum_{n=0}^{N-1} S^n \tilde{f}(\{\ell\}, T^{[\ell]} x)$ exists and is equal to $\mathbb{E}(\tilde{f} \mid \mathcal{I}(S))(\{\ell\}, T^{[\ell]} x)$ for μ-a.e. $x \in X$. This implies that the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T^{[\gamma n+\ell]} f(x)$$

exists for μ-a.e. $x \in X$, and hence its pointwise limit is equal to its $L^2(\mu)$ limit. So,

$$\mathbb{E}(\tilde{f} \mid \mathcal{I}(S))(\{\ell\}, T^{[\ell]} x) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} S^n \tilde{f}(\{\ell\}, T^{[\ell]} x) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T^{[\gamma n+\ell]} f(x)$$

for μ-a.e. $x \in X$. The result now follows by Proposition 4.4. □

5. *Proof of main results*

In this last section, we give the proof of the statements in §1. To lighten the notation, we omit writing the spaces along which we integrate, since they are easily deduced by the measures that we use.

Proof of Theorem 1.4. The case $d = 1$ is true by the definition of $\mathcal{S}^*$. Now suppose that the conclusion holds for $d - 1$. Let ν be a coupling of the systems (X_i, μ_i) and let $\pi_1 : X_1 \times \cdots \times X_d \to X_1$ be the projection to the first coordinate and $\pi_2 : X_1 \times \cdots \times X_d \to X_2 \times \cdots \times X_d$ be the projection to the rest of the coordinates. Write $\vec{x} = (x_1, x_*)$, where $x_* = (x_2, \ldots, x_d)$. By the induction hypothesis, $(\pi_1)_* \lambda_{\vec{x}} = \mu_{[T_1], x_1}$ ν-a.e. and $(\pi_2)_* \lambda_{\vec{x}} = \mu_{[T_2,\ldots,T_d], x_*}$ ν-a.e.

For every $1 \le i \le d$, fix a countable dense set of continuous functions $\mathcal{C}_i = \{g_{i,k} : k \in \mathbb{N}\} \subseteq C(X_i)$. Let $X_1' \subseteq X_1$ be a full μ_1-measure set such that $((1/N) \sum_{n=0}^{N-1} T_1^n g_{1,k}(x_1))_N$ converges to $\mathbb{E}(g_{1,k} \mid \mathcal{I}(T_1))(x_1) = \int g_{1,k}\, d\mu_{[T_1], x_1}$ as $N \to \infty$ for every $k \in \mathbb{N}$ and $x_1 \in X_1'$. Since $\mu_1 = \int \mu_{[T_1], x_1}\, d\mu_1(x_1)$, the same is true for $\mu_{[T_1], x_1}$-a.e. $y \in X_1$ and for μ_1-a.e. $x_1 \in X_1$. Let $f_i \in \mathcal{C}_i$, $1 \le i \le d$ and $\vec{x} \in \pi_1^{-1}(X_1')$. By Lemma 3.1 and since $(\pi_1)_* \lambda_{\vec{x}} =$

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

$\mu_{[T_1],x_1}$, we have that

$$\begin{aligned}\int f_1 \otimes f_2 \otimes \cdots \otimes f_d \, d\lambda_{\vec{x}} &= \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \int T_1^n f_1 \otimes f_2 \otimes \cdots \otimes f_d \, d\lambda_{\vec{x}} \\ &= \int \bigg(\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} T_1^n f_1 \bigg) \otimes f_2 \otimes \cdots \otimes f_d \, d\lambda_{\vec{x}} \\ &= \int \mathbb{E}(f_1 \mid \mathcal{I}(T_1))(x_1) \otimes f_2 \otimes \cdots \otimes f_d \, d\lambda_{\vec{x}} \\ &= \int \bigg(\int f_1 \, d\mu_{[T_1],x_1} \bigg) \otimes f_2 \otimes \cdots \otimes f_d \, d\lambda_{\vec{x}}.\end{aligned}$$

We remark that the function $(\int f_1 \, d\mu_{[T_1],x_1})$ is constant $\mu_{[T_1],x_1}$-a.e. and thus equal $\mu_{[T_1],x_1}$-a.e. to a continuous function. Using again the definition of $\lambda_{\vec{x}}$ and the induction hypothesis, we get that

$$\begin{aligned}\int f_1 \otimes f_2 \otimes \cdots \otimes f_d \, d\lambda_{\vec{x}} &= \bigg(\int f_1 \, d\mu_{[T_1],x_1}\bigg)\bigg(\int f_2 \otimes \cdots \otimes f_d \, d\mu_{[T_2,\ldots,T_d],x_*}\bigg) \\ &= \mathbb{E}(f_1 \mid \mathcal{I}(T_1))(x_1)\mathbb{E}(f_2 \mid \mathcal{I}(T_2))(x_2) \cdots \mathbb{E}(f_d \mid \mathcal{I}(T_d))(x_d).\end{aligned}$$

By the density of linear combinations of functions $f_1 \otimes \cdots \otimes f_d$, $f_i \in \mathcal{C}_i$ in $C(X_1 \times \cdots \times X_d)$, we get that $\lambda_{\vec{x}} = \mu_{[T_1,\ldots,T_d],\vec{x}}$. □

We are now ready to prove Theorem 1.1 via Theorem 1.4.

Proof of Theorem 1.1. We keep the notation as above and let $\mathcal{C}_i$ be a countable family of continuous functions that is dense in $L^2(\mu_i)$ for $1 \le i \le d$. For $N \in \mathbb{N}$, denote

$$A_N(x_1, \ldots, x_d) := \frac{1}{N}\sum_{n=0}^{N-1} f_1(T_1^{[a_1(n)]}x_1) \cdots f_d(T_d^{[a_d(n)]}x_d).$$

For $k \in \mathbb{N}$, pick $\widehat{f}_{i,k} \in \mathcal{C}_i$ such that $\|f_i - \widehat{f}_{i,k}\|_2 \le 1/k$ for $1 \le i \le d$ and denote

$$\widehat{A}_{N,k}(x_1, \ldots, x_d) := \frac{1}{N}\sum_{n=0}^{N-1} \widehat{f}_{1,k}(T_1^{[a_1(n)]}x_1) \cdots \widehat{f}_{d,k}(T_d^{[a_d(n)]}x_d).$$

By the definition of $\mathcal{S}^*$ and the telescoping inequality,

$$\limsup_{N\to\infty} |A_N(x_1, \ldots, x_d) - \widehat{A}_{N,k}(x_1, \ldots, x_d)| \le \sum_{i=1}^{d} \mathbb{E}(|f_i - \widehat{f}_{i,k}| \mid \mathcal{I}(T_i))(x_i)$$

for ν-a.e. $\vec{x} = (x_1, \ldots, x_d) \in X_1 \times \cdots \times X_d$ and for all $k \in \mathbb{N}$. So, for every $k \in \mathbb{N}$,

$$\limsup_{N\to\infty} \bigg| A_N(x_1, \ldots, x_d) - \int f_1 \otimes \cdots \otimes f_d \, d\mu_{[T_1,\ldots,T_d],\vec{x}} \bigg|$$

is bounded by the sum of the terms

$$\limsup_{N\to\infty} |A_N(x_1, \ldots, x_d) - \widehat{A}_{N,k}(x_1, \ldots, x_d)|; \tag{20}$$

$$\limsup_{N\to\infty} \bigg| \widehat{A}_{N,k}(x_1, \ldots, x_d) - \int \widehat{f}_{1,k} \otimes \cdots \otimes \widehat{f}_{d,k} \, d\mu_{[T_1,\ldots,T_d],\vec{x}} \bigg|; \tag{21}$$

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

and

$$\left| \int \widehat{f}_{1,k} \otimes \cdots \otimes \widehat{f}_{d,k} \, d\mu_{[T_1,\ldots,T_d],\vec{x}} - \int f_1 \otimes \cdots \otimes f_d \, d\mu_{[T_1,\ldots,T_d],\vec{x}} \right|. \tag{22}$$

By Theorem 1.4, term (21) is equal to zero. Again by telescoping, the sum of terms (20) and (22) is bounded by $2\sum_{i=1}^{d} \mathbb{E}(|f_i - \widehat{f}_{i,k}| \mid \mathcal{I}(T_i))(x_i)$, for ν-a.e. $(x_1, \ldots, x_d) \in X_1 \times \cdots \times X_d$. If B_m denotes the set of points $\vec{x} = (x_1, \ldots, x_d)$ such that

$$\limsup_{N\to\infty} \left| A_N(x_1, \ldots, x_d) - \int f_1 \otimes \cdots \otimes f_d \, d\mu_{[T_1,\ldots,T_d],\vec{x}} \right| \geq \frac{1}{m},$$

then Markov's inequality implies that, for every $m \in \mathbb{N}$, the measure of B_m is smaller than $2dm/k$. Since k is arbitrary, we have that $\nu(B_m) = 0$ and $\nu(\bigcap_{m\in\mathbb{N}} B_m^c) = 1$. It is immediate to check that $A_N(x_1, \ldots, x_d)$ converges for every $(x_1, \ldots, x_d) \in \bigcap_{m\in\mathbb{N}} B_m^c$. □

Proof of Theorem 1.5. Let $x \in X$ and $\lambda_{N,x} = (1/N)\sum_{n=0}^{N-1} T^{[a(n)]}\delta_x$. By Lemma 3.1, we have that any weak limit of $\lambda_{N,x}$ is T-invariant and hence equal to μ by unique ergodicity. Therefore, $\lambda_{N,x}$ converges to μ as $N \to \infty$ and the conclusion follows. □

Proof of Theorem 1.6. This is the same proof as of Theorem 1.4, combined with the fact that the single average converges by Theorem 1.5. □

Proof of Theorem 1.2. (i) *The case where $a_1(n) = pn$, $p \in \mathbb{Z}\backslash\{0\}$.* The proof is very similar to the one of Theorem 1.4; we sketch it for completeness. We will give an expression for any weak limit of the average of the Dirac measure $\delta_{\vec{x}}$, $\vec{x} = (x_1, \ldots, x_d)$ in a dense family of functions, and then it is routine to arrive at the conclusion (as in the proof of Theorem 1.1).

Let π_1 and π_2 be the projections as in the proof of Theorem 1.4. By Birkhoff's ergodic theorem, we have that $(\pi_1)_*\lambda_{\vec{x}} = \mu_{[T_1^p],x_1}$ for ν-a.e. $\vec{x} \in X_1 \times \cdots \times X_d$. By Theorem 1.4, $(\pi_2)_*\lambda_{\vec{x}} = \mu_{[T_2,\ldots,T_d],x_*}$ for ν-a.e. $\vec{x} \in X_1 \times \cdots \times X_d$.

Fix a countable dense set of continuous functions $\mathcal{C}_i = \{g_{i,k} : k \in \mathbb{N}\} \subseteq C(X_i)$. We have that there exists a set of full measure $X_1' \subseteq X_1$ such that $((1/N)\sum_{n=0}^{N-1} T_1^{pn} g_{1,k}(x_1))_N$ converges to $\mathbb{E}(g_{1,k} \mid \mathcal{I}(T_1^p))(x_1) = \int g_{1,k} \, d\mu_{[T_1^p],x_1}$ as $N \to \infty$ for every $k \in \mathbb{N}$ and $x_1 \in X_1'$. Since $\mu_1 = \int \mu_{[T_1^p],x_1} \, d\mu_1(x_1)$, the same is true for $\mu_{[T_1^p],x_1}$-a.e. $y \in X_1$ for μ_1-a.e. $x_1 \in X_1$. Let $f_i \in \mathcal{C}_i$, $1 \leq i \leq d$ and $\vec{x} \in \pi_1^{-1}(X_1')$. Applying Lemma 3.2 we have (as in Theorem 1.4)

$$\begin{aligned}\int f_1 \otimes f_2 \otimes \cdots \otimes f_d \, d\lambda_{\vec{x}} &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int T_1^{pn} f_1 \otimes f_2 \otimes \cdots \otimes f_d \, d\lambda_{\vec{x}} \\ &= \int \left(\int f_1 \, d\mu_{[T_1^p],x_1} \right) \otimes f_2 \otimes \cdots \otimes f_d \, d\lambda_{\vec{x}}.\end{aligned}$$

Remark that the function $(\int f_1 \, d\mu_{[T_1^p],x_1})$ is constant $\mu_{[T_1^p],x_1}$-a.e. and thus equal $\mu_{[T_1],x_1}$-a.e. to a continuous function. Using again the definition of $\lambda_{\vec{x}}$ and Theorem 1.4, we get that

$$\begin{aligned}\int f_1 \otimes f_2 \otimes \cdots \otimes f_d \, d\lambda_{\vec{x}} &= \left(\int f_1 \, d\mu_{[T_1^p],x_1} \right) \left(\int f_2 \otimes \cdots \otimes f_d \, d\mu_{[T_2,\ldots,T_d],x_*} \right) \\ &= \mathbb{E}(f_1 \mid \mathcal{I}(T_1^p))(x_1)\mathbb{E}(f_2 \mid \mathcal{I}(T_2))(x_2) \cdots \mathbb{E}(f_d \mid \mathcal{I}(T_d))(x_d).\end{aligned}$$

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

By the density of linear combinations of functions $f_1 \otimes \cdots \otimes f_d$, $f_i \in \mathcal{C}_i$ in $C(X_1 \times \cdots \times X_d)$, we get that $\lambda_{\vec{x}} = \mu_{[T_1^p, T_2, \ldots, T_d], \vec{x}}$.

(ii) *The case where* $a_1(n) = pn + \ell,\ p \in \mathbb{Z}\backslash\{0\}$. Using case (i),

$$\begin{aligned}
&\int f_1 \otimes f_2 \otimes \cdots \otimes f_d \, d\lambda_x \\
&\quad = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T_1^{[pn+\ell]} f_1(x_1) T_2^{[a_2(n)]} f_2(x_2) \cdots T_d^{[a_d(n)]} f_d(x_d) \\
&\quad = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T_1^{pn} (T_1^{[\ell]} f_1)(x_1) T_2^{[a_2(n)]} f_2(x_2) \cdots T_d^{[a_d(n)]} f_d(x_d) \\
&= \mathbb{E}(T_1^{[\ell]} f_1 \mid \mathcal{I}(T_1^p))(x_1) \mathbb{E}(f_2 \mid \mathcal{I}(T_2))(x_2) \cdots \mathbb{E}(f_d \mid \mathcal{I}(T_d))(x_d).
\end{aligned}$$

(iii) *The case where* $a_1(n) = kn + \ell, k = p/q \in \mathbb{Q}\backslash\{0\}$. Using case (ii),

$$\begin{aligned}
&\int f_1 \otimes f_2 \otimes \cdots \otimes f_d \, d\lambda_{\vec{x}} \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T_1^{[kn+\ell]} f_1(x_1) T_2^{[a_2(n)]} f_2(x_2) \cdots T_d^{[a_d(n)]} f_d(x_d) \\
&= \frac{1}{q} \sum_{j=0}^{q-1} \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T_1^{[k(qn+j)+\ell]} f_1(x_1) T_2^{[a_2(qn+j)]} f_2(x_2) \cdots T_d^{[a_d(qn+j)]} f_d(x_d) \\
&= \frac{1}{q} \sum_{j=0}^{q-1} \mathbb{E}(T_1^{[pj/q+\ell]} f_1 \mid \mathcal{I}(T_1^p))(x_1) \mathbb{E}(f_2 \mid \mathcal{I}(T_2))(x_2) \cdots \mathbb{E}(f_d \mid \mathcal{I}(T_d))(x_d).
\end{aligned}$$

(iv) *The case where* $a_1(n) = \gamma n + \ell,\ \gamma \in \mathbb{R}\backslash\mathbb{Q}$ (recall the notation of Corollary 4.5). By case (i), and by passing to the mapping torus extension,

$$\begin{aligned}
&\int f_1 \otimes f_2 \otimes \cdots \otimes f_d \, d\lambda_{\vec{x}} \\
&\quad = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T_1^{[\gamma n+\ell]} f_1(x_1) T_2^{[a_2(n)]} f_2(x_2) \cdots T_d^{[a_d(n)]} f_d(x_d) \\
&\quad = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} S^n \tilde{f}_1(\{\ell\}, T^{[\ell]} x_1) T_2^{[a_2(n)]} f_2(x_2) \cdots T_d^{[a_d(n)]} f_d(x_d) \\
&\quad = \mathbb{E}(\tilde{f}_1 \mid \mathcal{I}(S))(\{\ell\}, T^{[\ell]} x_1) \mathbb{E}(f_2 \mid \mathcal{I}(T_2))(x_2) \cdots \mathbb{E}(f_d \mid \mathcal{I}(T_d))(x_d).
\end{aligned}$$

By Corollary 4.5, the result follows. □

Proof of Corollary 1.3. This follows by Theorems 1.1 and 1.2. □

Acknowledgements. The first author is supported by Fondecyt Iniciación en Investigación Grant 11160061. We thank Nikos Frantzikinakis for bringing us to the alternative proof of Theorem 1.1 for functions of different growth from the class $\mathcal{LE}_\varepsilon$. We especially thank the anonymous referee whose wide knowledge of the topic allowed us to add important historical references and motivations to the problem we studied.

Appendix. Classes of functions used in the paper

In this appendix, we summarize all the classes of functions pertaining to this paper.

$\mathcal{F}$: the set of all Fejér functions.

$\mathcal{L}_*$: the set of linear functions with non-zero leading coefficient.

$\mathcal{H}$: the union of all Hardy fields.

$\mathcal{H}_\varepsilon$: the set of functions a that belongs to some Hardy field and satisfy $x^\varepsilon \prec a(x) \prec x$, for some $\varepsilon > 0$.

$\mathcal{LE}$: the set of logarithmico-exponential Hardy field functions.

$\mathcal{LE}_\varepsilon$: the set of logarithmico-exponential Hardy field functions a that satisfy the growth condition $x^\varepsilon \prec a(x) \prec x$ for some $\varepsilon > 0$.

$\mathcal{SL}$: the set of sublinear functions (i.e., functions a that satisfy $a(x) \prec x$).

$$\mathcal{R} = \left\{a \in \mathcal{C}^3(\mathbb{R}^+) : \text{the limits } \lim_{x\to\infty} \frac{xa'(x)}{a(x)}, \lim_{x\to\infty} \frac{xa''(x)}{a'(x)}, \right.$$
$$\left. \text{and } \lim_{x\to\infty} \frac{xa'''(x)}{a''(x)} \text{ exist in } \mathbb{R}\right\}.$$

$$\mathcal{T} = \left\{a \in \mathcal{R} : \lim_{x\to\infty} \frac{xa'(x)}{a(x)} \in (0, 1) \text{ or } \lim_{x\to\infty} \frac{xa'(x)}{a(x)} = 1 \right.$$
$$\left. \text{and } \lim_{x\to\infty} a'(x) = 0 \text{ monotonically}\right\}.$$

$$D_k = \left\{a : \limsup_{x\to\operatorname{sgn}(a^{-1})\cdot\infty} \sup_{h\in[-1,1]} \left|\frac{a^{(k+1)}(x+h)}{a^{(k)}(x)}\right| < \infty\right\}.$$

$$M_k = \{a : a^{(k)} \text{ is eventually monotone}\}.$$

$$S = \left\{a \in \mathcal{C}^3(\mathbb{R}^+) \;\middle|\; a, \frac{1}{a'} \in \mathcal{SL} \text{ and } a^{-1} \in M_1 \cap D_0 \cap D_1 \cap (D_2 \cup M_2)\right\}.$$

$\mathcal{S}^*$: the set of functions $a \in \mathcal{S}$, where $\lim_{N\to\infty} (1/N) \sum_{n=0}^{N-1} f(T^{[a(n)]}x)$ exists pointwisely (almost everywhere) for every measure-preserving system (X, μ, T) and every bounded measurable function f.

The relation between the classes of main interest is $\mathcal{H}_\varepsilon \subsetneq \mathcal{T} \subsetneq \mathcal{S}^* \subsetneq \mathcal{S} \subsetneq \mathcal{F} \subsetneq \mathcal{SL}$.

References

[1] I. Assani. Multiple recurrence and almost sure convergence of weakly mixing dynamical systems. *Israel J. Math.* **103** (1998), 111–125.

[2] I. Assani. Pointwise convergence of nonconventional averages. *Collect. Math.* **102**(2) (2005), 245–262.

[3] I. Assani. Averages along cubes for not necessarily commuting measure preserving transformations. *Contemp. Math.* **430** (2007), 1–19.

https://doi.org/10.1017/etds.2018.118 Published online by Cambridge University Press

[**4**] I. Assani. Pointwise convergence of ergodic averages along cubes. *J. Anal. Math.* **110** (2010), 241–269.
[**5**] T. Austin. On the norm convergence of nonconventional ergodic averages. *Ergod. Th. & Dynam. Sys.* **30**(2) (2010), 321–338.
[**6**] D. Berend. Joint ergodicity and mixing. *J. Anal. Math.* **45** (1985), 255–284.
[**7**] V. Bergelson. Weakly mixing PET. *Ergod. Th. & Dynam. Sys.* **7**(3) (1987), 337–349.
[**8**] V. Bergelson. *The Multifarious Poincaré Recurrence Theorem (Descriptive Set Theory and Dynamical Systems)*. Eds. M. Foreman, A. S. Kechris, A. Louveau and B. Weiss. Cambridge University Press, New York, 2000, pp. 31–57.
[**9**] V. Bergelson and I. Håland-Knutson. Weakly mixing implies mixing of higher orders along tempered functions. *Ergod. Th. & Dynam. Sys.* **29**(5) (2009), 1375–1416.
[**10**] V. Bergelson and A. Leibman. A nilpotent Roth theorem. *Invent. Math.* **147** (2002), 429–470.
[**11**] M. Boshernitzan, G. Kolesnik, A. Quass and M. Weirdl. Ergodic averaging sequences. *J. Anal. Math.* **95** (2005), 63–103.
[**12**] J. Bourgain. Pointwise ergodic theorems for arithmetic sets. *Publ. Math. Inst. Hautes Études Sci.* **69** (1989), 5–45.
[**13**] J. Bourgain. Double recurrence and almost sure convergence. *J. Reine Angew. Math.* **404** (1990), 140–161.
[**14**] Q. Chu. Convergence of multiple ergodic averages along cubes for several commuting transformations. *Studia Math.* **196**(1) (2009), 13–22.
[**15**] Q. Chu, N. Frantzikinakis and B. Host. Ergodic averages of commuting transformations with distinct degree polynomial iterates. *Proc. Lond. Math. Soc. (3)* **102** (2011), 801–842.
[**16**] Q. Chu and N. Frantzikinakis. Pointwise convergence for cubic and polynomial ergodic averages of non-commuting transformations. *Ergod. Th. & Dynam. Sys.* **32**(3) (2012), 877–897.
[**17**] S. Donoso and W. Sun. A pointwise cubic average for two commuting transformations. *Israel J. Math.* **216**(2) (2016), 657–678.
[**18**] S. Donoso and W. Sun. Pointwise multiple averages for systems with two commuting transformations. *Ergod. Th. & Dynam. Sys.* **38**(6) (2018), 2132–2157.
[**19**] S. Donoso and W. Sun. Pointwise convergence of some multiple ergodic averages. *Adv. Math.* **330** (2018), 946–996.
[**20**] N. Frantzikinakis. Multiple recurrence and convergence for Hardy field sequences of polynomial growth. *J. Anal. Math.* **112** (2010), 79–135.
[**21**] N. Frantzikinakis. A multidimensional Szemeredi theorem for Hardy sequences of polynomial growth. *Trans. Amer. Math. Soc.* **367**(8) (2015), 5653–5692.
[**22**] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Anal. Math.* **31** (1977), 204–256.
[**23**] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.* **11** (2001), 465–588.
[**24**] Y. Gutman, W. Huang, S. Shao and X. Ye. Almost sure convergence of the multiple ergodic average for certain weakly mixing systems. *Acta Math. Sin. (Engl. Ser.)* **34**(1) (2018), 79–90.
[**25**] B. Host. Ergodic seminorms for commuting transformations and applications. *Studia Math.* **195**(1) (2009), 31–49.
[**26**] B. Host and B. Kra. Nonconventional averages and nilmanifolds. *Ann. of Math. (2)* **161**(1) (2005), 398–488.
[**27**] W. Huang, S. Shao and X. Ye. Pointwise convergence of multiple ergodic averages and strictly ergodic models. *J. Anal. Math.* (to appear). *Preprint*, 2014, arXiv:1406.5930.
[**28**] W. Huang, S. Shao and X. Ye. Strictly ergodic models under face and parallelepiped group actions. *Commun. Math. Stat.* **5**(1) (2017), 93–122.
[**29**] A. Koutsogiannis. Integer part polynomial correlation sequences. *Ergod. Th. & Dynam. Sys.* **38**(4) (2018), 1525–1542.
[**30**] A. Koutsogiannis. Closest integer polynomial multiple recurrence along shifted primes. *Ergod. Th. & Dynam. Sys.* **38**(2) (2018), 666–685.
[**31**] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Ergod. Th. & Dynam. Sys.* **28**(2) (2008), 657–688.
[**32**] M. Walsh. Norm convergence of nilpotent ergodic averages. *Ann. of Math. (2)* **175**(3) (2012), 1667–1688.