Ergod. Th. & Dynam. Sys., (2022), **42**, 2270–2280 © The Author(s), 2021. Published by Cambridge University Press.
doi:10.1017/etds.2021.45

Pointwise convergence of certain continuous-time double ergodic averages

MICHAEL CHRIST†, POLONA DURCIK‡, VJEKOSLAV KOVAČ§ and JORIS ROOS¶,‖

† *Department of Mathematics, University of California, Berkeley, CA 94720, USA*
(e-mail: mchrist@berkeley.edu)
‡ *Schmid College of Science and Technology, Chapman University, One University Drive, Orange, CA 92866, USA*
(e-mail: durcik@chapman.edu)
§ *Department of Mathematics, Faculty of Science, University of Zagreb, 10000 Zagreb, Croatia*
(e-mail: vjekovac@math.hr)
¶ *Department of Mathematical Sciences, University of Massachusetts Lowell, Lowell, MA 01854, USA*
‖ *School of Mathematics, The University of Edinburgh, Edinburgh, EH9 3FD, UK*
(e-mail: joris_roos@uml.edu)

(*Received* 13 *November* 2020 *and accepted in revised form* 19 *March* 2021)

Abstract. We prove almost everywhere convergence of continuous-time quadratic averages with respect to two commuting $\mathbb{R}$-actions, coming from a single jointly measurable measure-preserving $\mathbb{R}^2$-action on a probability space. The key ingredient of the proof comes from recent work on multilinear singular integrals; more specifically, from the study of a curved model for the triangular Hilbert transform.

Key words: multiple ergodic average, convergence almost everywhere, Calderón transference principle, multilinear estimate
2020 Mathematics Subject Classification: 37A30 (Primary); 37A46 (Secondary)

1. *Introduction*

In this article, we apply recent progress in multilinear harmonic analysis [11, 12] to a problem on convergence almost everywhere (a.e.) in the ergodic theory.

Suppose there is an action of the group $\mathbb{R}^2$ on a probability space $(X, \mathcal{F}, \mu)$,

$$\mathbb{R}^2 \times X \to X, \quad (g, x) \mapsto g \cdot x,$$

which is jointly measurable and measure preserving. In the language of Varadarajan [25], $(X, \mathcal{F})$ is a Borel $\mathbb{R}^2$-space and μ is an invariant measure.

An alternative way of looking at this set-up is to define mutually commuting one-parameter groups of $(\mathcal{F}, \mathcal{F})$-measurable measure-$\mu$-preserving transformations $(S^t : X \to X)_{t\in\mathbb{R}}$ and $(T^t : X \to X)_{t\in\mathbb{R}}$ by

$$S^t x := (t, 0) \cdot x, \quad T^t x := (0, t) \cdot x$$

for every $t \in \mathbb{R}$ and $x \in X$. That way, the above $\mathbb{R}^2$-action can be rewritten simply as $((s, t), x) \mapsto S^s T^t x$, but note that we also require joint measurability of this map. On the other hand, $(t, x) \mapsto S^t x$ and $(t, x) \mapsto T^t x$ are two mutually commuting measure-preserving $\mathbb{R}$-actions, that is, flows. We find the latter viewpoint and notation more suggestive, as they emphasize analogies with the corresponding discrete set-up, that is, $\mathbb{Z}^2$-actions, which are determined simply by two commuting transformations $S = S^1$ and $T = T^1$; for example, see (1.2) and (1.3) below.

Fix $p, q \in [1, \infty]$ such that $1/p + 1/q \le 1$. We are interested in the continuous-time double averages

$$A_N(f_1, f_2)(x) := \frac{1}{N} \int_0^N f_1(S^t x) f_2(T^{t^2} x)\, dt, \tag{1.1}$$

defined for a positive real number N, functions $f_1 \in \mathrm{L}^p(X)$ and $f_2 \in \mathrm{L}^q(X)$, and a point $x \in X$. If f_1 and f_2 are given, then, for μ-almost every x, the integrals in (1.1) exist and continuously depend on $N \in (0, \infty)$. Indeed, the Tonelli–Fubini theorem, Hölder's inequality, monotonicity of the $\mathrm{L}^p(X)$-norms and the fact that S^t, T^{t^2} preserve measure μ, together, imply that

$$\int_X \int_0^M |f_1(S^t x) f_2(T^{t^2} x)|\, dt\, d\mu(x) \le M \|f_1\|_{\mathrm{L}^p(X)} \|f_2\|_{\mathrm{L}^q(X)} < \infty$$

for any positive number M. Most of the literature that studies multiple ergodic averages simply takes the functions to be in $\mathrm{L}^\infty(X)$.

General single-parameter polynomial multiple ergodic averages were introduced by Bergelson and Leibman [3, 4], albeit in a discrete setting. The averages (1.1) constitute the simplest case of such polynomial (but not purely linear) averages with respect to several commuting group actions. This article establishes their convergence a.e.

THEOREM 1.1. *Let $((s, t), x) \mapsto S^s T^t x$ be a jointly measurable measure-preserving action of $\mathbb{R}^2$ on a probability space $(X, \mathcal{F}, \mu)$. Let $p, q \in (1, \infty]$ satisfy $1/p + 1/q \le 1$. Let $f_1 \in \mathrm{L}^p(X)$ and $f_2 \in \mathrm{L}^q(X)$. Then, for μ-almost every $x \in X$, the limit*

$$\lim_{N\to\infty} \frac{1}{N} \int_0^N f_1(S^t x) f_2(T^{t^2} x)\, dt$$

exists.

https://doi.org/10.1017/etds.2021.45 Published online by Cambridge University Press

To the authors' knowledge, this is the first result on pointwise convergence of some single-parameter multiple ergodic averages with respect to two general commuting $\mathbb{R}$-actions, without any structural assumptions on the measure-preserving system in question.

Generalizations of continuous-time single-parameter averages (1.1) to $\mathbb{R}^D$-actions, several polynomials, and several functions were studied by Austin [2]. He showed that these multiple averages always converge in the L^2-norm when the functions are taken from $L^\infty(X)$. The paper [2] also emphasizes simplifications coming from working in the continuous-time setting, as opposed to the discrete one. The most notable simplification comes from the ability to change variables in integrals with respect to the time-variable. Bergelson, Leibman, and Moreira [5] went a step further by giving general principles for deducing continuous results on convergence of various ergodic averages from their discrete analogues. A discrete-time analogue of Austin's L^2-convergence result was later established (in the greater generality of nilpotent group actions) by Walsh [26].

However, pointwise results on single-parameter multiple ergodic averages are much more difficult in either of the two settings. Without further structural assumptions, a.e. convergence is only known for double averages with respect to a single (invertible bi-measurable) measure-preserving transformation $T\colon X \to X$,

$$\frac{1}{N}\sum_{n=0}^{N-1} f_1(T^{P_1(n)}x) f_2(T^{P_2(n)}x),$$

when either P_1, P_2 are both linear polynomials (a result by Bourgain [6], with its continuous-time analogue formulated explicitly in [5, Theorem 8.30]) or when P_1 is linear and P_2 has degree greater than one (a recent result by Krause, Mirek, and Tao [22]). The latter case naturally motivates the study of averages

$$\frac{1}{N}\sum_{n=0}^{N-1} f_1(S^n x) f_2(T^{n^2} x), \tag{1.2}$$

where $S, T\colon X \to X$ are now two commuting (invertible bi-measurable) measure-preserving transformations. Convergence a.e. of (1.2) is still open at the time of writing and Theorem 1.1 solves a continuous-time analogue of this problem. As yet another source of motivation, we mention that a.e. convergence of purely linear double averages

$$\frac{1}{N}\sum_{n=0}^{N-1} f_1(S^n x) f_2(T^n x) \tag{1.3}$$

is also a well-known open problem for general commuting S and T; see the survey paper by Frantzikinakis [18]. On the other hand, continuous-time analogues of (1.3) are thought to be equally as difficult as (1.3) themselves: crucial differences disappear in the case of linear powers of transformations. We remark, in passing, that a.e. convergence is known for various multi-parameter multiple ergodic averages, such as two types of 'cubic' averages (see [1, 13–15]) or 'additionally averaged' averages (see [15, 16, 20]). Questions on convergence of such averages tend to be easier, but these objects appear naturally in studies of single-parameter averages.

It may be of interest to establish more quantitative variants of Theorem 1.1. We exploit two non-quantitative reductions. We use a maximal function inequality combined with convergence on a dense subset (as opposed to bounding a certain variational norm, as in [**7**, **8**, **22**]), and we work with lacunary sequences of scales (as opposed to discussing long and short jumps separately, as in [**21**]).

A minor modification of the proof presented here can establish a.e. convergence of variants of the averages (1.1) in which t^2 is replaced by t^κ for some fixed positive number $\kappa \neq 1$. Indeed, for the main technical ingredient of the proof, Theorem 1.1, this generalization is sketched in [**12**]. The particular choice $\kappa = 2$ is also used below in connection with (2.8) and (1.1), but, at those junctures of the proof, the restriction to $\kappa = 2$ is an inessential matter of convenience.

Let us also mention a vast generalization of Theorem 1.1 announced after this article was completed. Frantzikinakis [**17**, Theorem 1.9] used spectral techniques to show a.e. convergence of continuous-time multiple ergodic averages for (not necessarily commuting) $\mathbb{R}$-actions with functions of (not necessarily polynomial) 'different but not too different' growth in t in the exponents.

The rest of the paper is dedicated to the proof of Theorem 1.1. We can assume that $p, q \in (1, \infty)$ and $1/p + 1/q = 1$. Indeed, the L^p-spaces with respect to a finite measure are nested, which allows raising of either of the two exponents. Otherwise, the largest range of $(p, q) \in [1, \infty]^2$ in which the a.e. convergence result holds is not clear and even justification of the defining formula (1.1) is not immediate. A non-trivial L^1 counterexample for single-function discrete-time quadratic averages was given by Buczolich and Mauldin [**9**]; also, see [**23**] for an extension of their result.

1.1. *Notation.* For two functions $A, B \colon X \to [0, \infty)$ and a set of parameters P we write $A(x) \lesssim_P B(x)$ if the inequality $A(x) \leq C_P B(x)$ holds for each $x \in X$ with a constant C_P depending on the parameters from P, but independent of x. Let $\mathbb{1}_S$ denote the *indicator function* of a set $S \subseteq X$, where the ambient set X is understood from the context. The *floor* of $x \in \mathbb{R}$ will be denoted by $\lfloor x \rfloor$; it is the largest integer not exceeding x.

If $(X, \mathcal{F}, \mu)$ is a measure space and $p \in [1, \infty)$, then the L^p*-norm* of an $\mathcal{F}$-measurable function $f \colon X \to \mathbb{C}$ is defined as

$$\|f\|_{\mathrm{L}^p(X)} := \bigg(\int_X |f(x)|^p \, d\mu(x) \bigg)^{1/p}.$$

We also set

$$\|f\|_{\mathrm{L}^\infty(X)} := \operatorname*{ess\,sup}_{x \in X} |f(x)|.$$

On the other hand, the *weak* L^p*-norm* is defined as

$$\|f\|_{\mathrm{L}^{p,\infty}(X)} := \Big(\sup_{\alpha \in (0,\infty)} \alpha^p \mu(\{x \in X : |f(x)| > \alpha\}) \Big)^{1/p}.$$

Occasionally, the variable with respect to which the norm is taken will be denoted in the subscript, so that we can write $\|f(x)\|_{\mathrm{L}^p_x(X)}$ in place of $\|f\|_{\mathrm{L}^p(X)}$. On $\mathbb{R}^d$, the Lebesgue measure will always be understood.

The *Fourier transform* of $f \in \mathrm{L}^1(\mathbb{R}^d)$ is defined as

$$\widehat{f}(\xi) := \int_{\mathbb{R}^d} f(x) e^{-2\pi i x \cdot \xi}\, dx$$

for each $\xi \in \mathbb{R}^d$, where $(x, y) \mapsto x \cdot y$ is the standard scalar product on $\mathbb{R}^d$. The map $f \mapsto \widehat{f}$ extends by continuity to the space $\mathrm{L}^2(\mathbb{R}^d)$, where it becomes a linear isometric isomorphism.

We write $\mathrm{span}(S)$ for the linear span of a set of vectors S in some linear space. If V and W are mutually orthogonal subspaces of some inner product space, then $V \oplus W$ will denote their *(orthogonal) sum*, that is, the linear span of their union. Finally, $\mathrm{img}(L)$ and $\ker(L)$ will, respectively, denote the range and the null space of a linear operator L.

2. *Ergodic theory reductions*

Theorem 1.1 will be deduced from the following proposition that deals with functions on the real line.

PROPOSITION 2.1. *For each $\delta \in (0, 1]$, there exists a constant $\gamma \in (0, 1)$ such that*

$$\Big\| \frac{1}{N} \int_0^N (F_1(u+t+\delta, v) - F_1(u+t, v)) F_2(u, v+t^2)\, dt \Big\|_{\mathrm{L}^1_{(u,v)}(\mathbb{R}^2)} \lesssim_{\gamma,\delta} N^{-\gamma} \|F_1\|_{\mathrm{L}^2(\mathbb{R}^2)} \|F_2\|_{\mathrm{L}^2(\mathbb{R}^2)} \tag{2.1}$$

for every $N \in [1, \infty)$ and all $F_1, F_2 \in \mathrm{L}^2(\mathbb{R}^2)$.

The proof of Proposition 2.1 will be postponed until the next section. Moreover, we will see that the quantifiers can be reversed: we will be able to choose γ that works for each δ. Here we show how (2.1) implies the main result.

Proof of Theorem 1.1. Let $p^{-1} + q^{-1} = 1$. We begin by applying a variant of the so-called *lacunary subsequence trick*; see [**19**, Appendix A]. It reduces Theorem 1.1 to proving that

$$(A_{\alpha^n}(f_1, f_2)(x))_{n=0}^{\infty} \text{ converges in } \mathbb{C} \text{ for almost every } x \in X \tag{2.2}$$

for every fixed $\alpha \in (1, \infty)$. Indeed, we can assume that f_1 and f_2 are non-negative functions because, otherwise, we can split them, first into real and imaginary, and then into positive and negative parts. Denoting by $\lfloor y \rfloor$ the largest integer not exceeding a real number y, we can estimate

$$\alpha^{-1} A_{\alpha^{\lfloor \log_\alpha N \rfloor}}(f_1, f_2)(x) \le A_N(f, g)(x) \le \alpha A_{\alpha^{\lfloor \log_\alpha N \rfloor + 1}}(f_1, f_2)(x)$$

and this implies that

$$\begin{aligned} \alpha^{-1} \liminf_{\mathbb{N} \ni n \to \infty} A_{\alpha^n}(f_1, f_2)(x) &\le \liminf_{\mathbb{R} \ni N \to \infty} A_N(f_1, f_2)(x) \\ &\le \limsup_{\mathbb{R} \ni N \to \infty} A_N(f_1, f_2)(x) \le \alpha \limsup_{\mathbb{N} \ni n \to \infty} A_{\alpha^n}(f_1, f_2)(x). \end{aligned} \tag{2.3}$$

https://doi.org/10.1017/etds.2021.45 Published online by Cambridge University Press

By (2.2) applied with $\alpha = 2^{2^{-m}}$, we know that, at almost every point $x \in X$, the limit

$$\lim_{n\to\infty} A_{2^{n2^{-m}}}(f_1, f_2)(x)$$

exists for each positive integer m. Its value is independent of m, since the corresponding sequences are subsequences of each other, so we can denote it by $L(x) \in [0, \infty)$. For any such x, the estimate (2.3) gives

$$2^{-2^{-m}} L(x) \le \liminf_{N\to\infty} A_N(f_1, f_2)(x) \le \limsup_{N\to\infty} A_N(f_1, f_2)(x) \le 2^{2^{-m}} L(x),$$

so we may let $m \to \infty$ and conclude that $\lim_{N\to\infty} A_N(f_1, f_2)(x)$ exists and also equals $L(x)$.

We will also use the easy weak-type inequality

$$\Big\| \sup_{N\in(0,\infty)} |A_N(f_1, f_2)| \Big\|_{\mathrm{L}^{1,\infty}(X)} \lesssim_{p,q} \|f_1\|_{\mathrm{L}^p(X)} \|f_2\|_{\mathrm{L}^q(X)} \tag{2.4}$$

for every $N \in (0, \infty)$, $f_1 \in \mathrm{L}^p(X)$ and $f_2 \in \mathrm{L}^q(X)$. It will enable us to restrict attention to dense subspaces of functions $f_1 \in \mathrm{L}^p(X)$ and $f_2 \in \mathrm{L}^q(X)$ by the aforementioned a.e. convergence paradigm. In order to prove (2.4), one can first apply Hölder's inequality, followed by the change of variables $s = t^2$ and a dyadic splitting of the integral in the second term: that is,

$$|A_N(f_1, f_2)| \le \bigg(\frac{1}{N} \int_0^N |f_1(S^t x)|^p \, dt \bigg)^{1/p} \times \bigg(\sum_{m=1}^{\infty} 2^{-m/2} \frac{1}{2^{-m+1}N^2} \int_0^{2^{-m+1}N^2} |f_2(T^s x)|^q \, ds \bigg)^{1/q}.$$

Then one can take the supremum in N and recall Hölder's inequality in Lorentz spaces [**24**] to bound the left-hand side of (2.4) by

$$\Big\| \sup_{N\in(0,\infty)} \frac{1}{N} \int_0^N |f_1(S^t x)|^p \, dt \Big\|_{\mathrm{L}^{1,\infty}(X)}^{1/p} \Big\| \sup_{N\in(0,\infty)} \frac{1}{N} \int_0^N |f_2(T^t x)|^q \, dt \Big\|_{\mathrm{L}^{1,\infty}(X)}^{1/q}.$$

It remains to apply the maximal ergodic weak L^1 inequality to the functions $|f_1|^p$ and $|f_2|^q$. If one only wants to use the well-known discrete-time maximal ergodic theorem, one can borrow a trick from [**5**], that is, restrict the values of N to the grid $\delta\mathbb{Z}$ for some $\delta > 0$ and apply the discrete-time theory to the L^1 functions

$$g_1(x) := \frac{1}{\delta} \int_0^\delta |f_1(S^t x)|^p \, dt, \quad g_2(x) := \frac{1}{\delta} \int_0^\delta |f_2(T^t x)|^q \, dt.$$

This completes the proof of (2.4).

A strengthening of (2.4) with the ordinary (strong) L^1-norm on the left-hand side can be deduced by the method of transference from [**12**, Theorem 2], which deals with functions on the real line. We do not need this strengthening here, since weak-type maximal inequalities are sufficient for the intended purpose of extending a.e. convergence.

A crucial ingredient of the proof of Theorem 1.1 is the following estimate.

LEMMA 2.2. *For each $\delta \in (0, 1]$, there exists a constant $\gamma \in (0, 1]$ such that*

$$\Big\| \frac{1}{N} \int_0^N (f_1(S^{t+\delta} x) - f_1(S^t x)) f_2(T^{t^2} x)\, dt \Big\|_{\mathrm{L}^1_x(X)} \lesssim_{\gamma,\delta} N^{-\gamma} \|f_1\|_{\mathrm{L}^2(X)} \|f_2\|_{\mathrm{L}^2(X)} \quad (2.5)$$

for every $N \in [1, \infty)$ and every $f_1, f_2 \in \mathrm{L}^2(X)$.

Proof. We deduce (2.5) from Proposition 2.1 using the *Calderón transference principle* [**10**]. By homogeneity, it is sufficient to prove inequality (2.5) for functions f_1 and f_2 normalized to satisfy

$$\|f_1\|_{\mathrm{L}^2(X)} = \|f_2\|_{\mathrm{L}^2(X)} = 1.$$

For each $x \in X$ and $N \geq 1$, define functions $F_1^{x,N}, F_2^{x,N} : \mathbb{R}^2 \to \mathbb{C}$ by

$$F_j^{x,N}(u, v) := f_j(S^u T^v x) \mathbb{1}_{[0,3N]}(u) \mathbb{1}_{[0,2N^2]}(v)$$

for $(u, v) \in \mathbb{R}^2$ and $j = 1, 2$. Since the measure μ is invariant under the $\mathbb{R}^2$-action in question, we can rewrite the left-hand side of (2.5) as

$$\frac{1}{N^3} \int_0^N \int_0^{N^2} \int_X \Big| \frac{1}{N} \int_0^N (f_1(S^{t+\delta} S^u T^v x) - f_1(S^t S^u T^v x)) f_2(T^{t^2} S^u T^v x)\, dt \Big| \, d\mu(x)\, du\, dv$$

$$\leq \frac{1}{N^3} \int_X \Big\| \frac{1}{N} \int_0^N (F_1^{x,N}(u + t + \delta, v) - F_1^{x,N}(u + t, v)) F_2^{x,N}(u, v + t^2)\, dt \Big\|_{\mathrm{L}^1_{(u,v)}(\mathbb{R}^2)} \, d\mu(x).$$

An application of (2.1) with functions $F_1^{x,N}, F_2^{x,N}$ for each fixed $x \in X$ bounds the last display by a constant multiple of

$$\begin{aligned} &\frac{1}{N^3} \int_X N^{-\gamma} \frac{1}{2} (\|F_1^{x,N}\|^2_{\mathrm{L}^2(\mathbb{R}^2)} + \|F_2^{x,N}\|^2_{\mathrm{L}^2(\mathbb{R}^2)})\, d\mu(x) \\ &= N^{-\gamma} \frac{1}{N^3} \int_0^{3N} \int_0^{2N^2} \int_X \frac{1}{2} (|f_1(S^u T^v x)|^2 + |f_2(S^u T^v x)|^2)\, d\mu(x)\, du\, dv \\ &= 6N^{-\gamma} \tfrac{1}{2} (\|f_1\|^2_{\mathrm{L}^2(X)} + \|f_2\|^2_{\mathrm{L}^2(X)}) = 6N^{-\gamma}, \end{aligned}$$

where we have again used the invariance of μ. This completes the proof of (2.5). □

For each $t \in \mathbb{R}$, let U^t denote the unitary operator on $\mathrm{L}^2(X)$ given by the formula $U^t f := f \circ S^t$. Our final auxiliary claim is that

$$\mathrm{span} \Big(\bigcup_{\delta \in (0,1]} \mathrm{img}(U^\delta - I) \Big) \oplus \Big(\bigcap_{\delta \in (0,1]} \ker(U^\delta - I) \Big) \quad (2.6)$$

is a dense subspace of $\mathrm{L}^2(X)$. Indeed, this easily follows from $\mathrm{img}(U^\delta - I)^\perp = \ker(U^\delta - I)$ for each δ, which, in turn, is a consequence of the fact that $U^\delta - I$ is a normal operator.

We are now ready to complete the proof of Theorem 1.1. By the initial reduction and the maximal inequality (2.4), we need only establish (2.2) for each fixed $\alpha \in (1, \infty)$ and for functions $f_1, f_2 \in \mathrm{L}^2(X)$. The reason is, of course, that $\mathrm{L}^p(X) \cap \mathrm{L}^2(X)$ is dense in $\mathrm{L}^p(X)$, while $\mathrm{L}^q(X) \cap \mathrm{L}^2(X)$ is dense in $\mathrm{L}^q(X)$. By yet another application of (2.4), this time with $p = q = 2$, we see that it suffices to take f_1 from the dense subspace (2.6) of $\mathrm{L}^2(X)$. In other words, we can assume that f_1 is of the form

$$\sum_{k=1}^{m} (g_k \circ S^{\delta_k} - g_k) + h,$$

where $m \in \mathbb{N}$, $\delta_1, \ldots, \delta_m \in (0, 1]$, $g_1, \ldots, g_m, h \in \mathrm{L}^2(X)$, and h is such that $h \circ S^t = h$ for each $t \in (0, 1]$ and thus also for each $t \in [0, \infty)$. That way, the theorem is reduced to showing that, for any $f_1, f_2 \in \mathrm{L}^2(X)$ and any parameters $\alpha > 1$ and $\delta \in (0, 1]$, the two sequential limits

$$\lim_{n\to\infty} \frac{1}{\alpha^n} \int_0^{\alpha^n} (f_1(S^{t+\delta}x) - f_1(S^t x)) f_2(T^{t^2} x)\, dt \tag{2.7}$$

and

$$\lim_{n\to\infty} \frac{1}{\alpha^n} \int_0^{\alpha^n} f_2(T^{t^2} x)\, dt \tag{2.8}$$

exist (in $\mathbb{C}$) for almost every $x \in X$.

Estimate (2.5), applied with $N = \alpha^n$, and summation in n give

$$\int_X \sum_{n=0}^{\infty} \left| \frac{1}{\alpha^n} \int_0^{\alpha^n} (f_1(S^{t+\delta}x) - f_1(S^t x)) f_2(T^{t^2} x)\, dt \right| d\mu(x)$$
$$\lesssim_{\gamma,\delta} \sum_{n=0}^{\infty} \alpha^{-\gamma n} \|f_1\|_{\mathrm{L}^2(X)} \|f_2\|_{\mathrm{L}^2(X)} < \infty.$$

Thus, for almost every $x \in X$, the sequence in (2.7) converges to zero, as a general term of a convergent series.

The limit in (2.8) exists for almost every $x \in X$ by [**5**, Theorem 8.31], which claims the same for general polynomial averages of a single L^2 function and constitutes a continuous-time analogue of Bourgain's result from [**7**].

3. *Harmonic analysis reductions*

Proof of Proposition 2.1. Let ζ be a C^∞ function compactly supported in $\mathbb{R}^2 \times (\mathbb{R} \setminus \{0\})$.

THEOREM 1.1. [**12**] *There exist $C, \sigma > 0$ with the following property. Let $F_1, F_2 \in \mathrm{L}^2(\mathbb{R}^2)$ and let $\lambda \geq 1$. Suppose that, for at least one of the indices $j = 1, 2$, $\widehat{F}_j(\xi_1, \xi_2)$ vanishes whenever $|\xi_j| < \lambda$. Then*

$$\left\| \int_{\mathbb{R}} F_1(x + t, y)\, F_2(x, y + t^2)\, \zeta(x, y, t)\, dt \right\|_{\mathrm{L}^1_{(x,y)}(\mathbb{R}^2)} \leq C\lambda^{-\sigma} \|F_1\|_{\mathrm{L}^2(\mathbb{R}^2)} \|F_2\|_{\mathrm{L}^2(\mathbb{R}^2)}.$$

https://doi.org/10.1017/etds.2021.45 Published online by Cambridge University Press

For an auxiliary function ζ as before, any $\delta \in (0, 1]$ and any $F_1, F_2 \in \mathrm{L}^2(\mathbb{R}^2)$ define

$$B_\delta(F_1, F_2)(x, y) := \int_{\mathbb{R}} (F_1(x + t + \delta, y) - F_1(x + t, y))\, F_2(x, y + t^2)\, \zeta(x, y, t)\, dt. \tag{1.1}$$

We claim that, to prove Proposition 2.1, it suffices to prove that there exists $\gamma \in (0, 1)$ such that

$$\|B_\delta(F_1, F_2)\|_{\mathrm{L}^1(\mathbb{R}^2)} \leq C_{\gamma,\zeta}\delta^\gamma \|F_1\|_{\mathrm{L}^2(\mathbb{R}^2)}\|F_2\|_{\mathrm{L}^2(\mathbb{R}^2)} \tag{1.2}$$

for every $\delta \in (0, 1]$, for all $F_1, F_2 \in \mathrm{L}^2(\mathbb{R}^2)$, where $C_{\gamma,\zeta}$ is a constant depending on γ and ζ.

This is a standard reduction, but some care needs to be taken due to the minus sign appearing in $B_\delta(F_1, F_2)$. By using the equality

$$\frac{1}{N}\mathbb{1}_{(0,N]} = \sum_{k=1}^{\infty} 2^{-k}\frac{1}{2^{-k}N}\mathbb{1}_{(2^{-k}N, 2^{-k+1}N]}$$

and rescaling

$$F_j(x, y) \mapsto (2^{-k}N)^{3/2}F_j(2^{-k}Nx, (2^{-k}N)^2 y), \quad \delta \mapsto (2^{-k}N)^{-1}\delta,$$

inequality (2.1) follows if we can show existence of $\gamma \in (0, 1)$ such that

$$\begin{aligned}\bigg\|\int_1^2 (F_1(x+t+\delta, y) - F_1(x+t, y))F_2(x, y+t^2)\, dt\bigg\|_{\mathrm{L}^1_{(x,y)}(\mathbb{R}^2)} \\ \lesssim_\gamma \delta^\gamma \|F_1\|_{\mathrm{L}^2(\mathbb{R}^2)}\|F_2\|_{\mathrm{L}^2(\mathbb{R}^2)}\end{aligned} \tag{1.3}$$

for all $\delta > 0$. Since (1.3) is trivial for $\delta > 1$ by the Cauchy–Schwarz inequality, we can again assume that $\delta \in (0, 1]$. Next, let η be a smooth non-negative function supported in $[-1, 1]^2$ and such that $\sum_{m\in\mathbb{Z}^2}\eta_m = 1$, where $\eta_m(x, y) := \eta((x, y) - m)$ for all $(x, y) \in \mathbb{R}^2$. The left-hand side of (1.3) is majorized by

$$\begin{aligned}\sum_{m\in\mathbb{Z}^2}\bigg\|\int_1^2 ((\widetilde{\eta}_m F_1)(x+t+\delta, y) - (\widetilde{\eta}_m F_1)(x+t, y)) \\ \cdot (\widetilde{\eta}_m F_2)(x, y+t^2)\eta_m(x, y)\, dt\bigg\|_{\mathrm{L}^1_{(x,y)}(\mathbb{R}^2)},\end{aligned}$$

where $\widetilde{\eta}$ is a smooth non-negative function compactly supported in $[-20, 20]^2$, equal to 1 on $[-10, 10]^2$ and $\widetilde{\eta}_m(x, y) := \widetilde{\eta}((x, y) - m)$. To apply (1.2), we also need to pass to a smooth cut-off function in the t-variable. To this end, choose a smooth non-negative function φ compactly supported in $[1, 2]$ so that $\|\varphi - \mathbb{1}_{[1,2]}\|_{\mathrm{L}^1(\mathbb{R})} \leq \delta$. Applying (1.2) with $\zeta(x, y, t) = \eta(x, y)\varphi(t)$ and majorizing the error term by the Minkowski and Cauchy–Schwarz inequalities shows that the previous display is majorized by

$$(C_{\gamma,\zeta}\delta^\gamma + \delta)\sum_{m\in\mathbb{Z}^2}\|\widetilde{\eta}_m F_1\|_{\mathrm{L}^2(\mathbb{R}^2)}\|\widetilde{\eta}_m F_2\|_{\mathrm{L}^2(\mathbb{R}^2)}.$$

https://doi.org/10.1017/etds.2021.45 Published online by Cambridge University Press

By the Cauchy–Schwarz inequality for the sum in m, the previous display is at most a constant multiple of $\delta^\gamma \|F_1\|_{\mathrm{L}^2(\mathbb{R}^2)} \|F_2\|_{\mathrm{L}^2(\mathbb{R}^2)}$, which proves the claim, that is, it establishes Proposition 2.1, modulo the proof of (1.2).

Proof of (1.2). Let $R \geq 1$ be determined later. Decompose

$$F_1 = F_{1,R} + G_{1,R},$$

where $F_{1,R}$ is defined via its Fourier transform as

$$\widehat{F_{1,R}}(\xi_1, \xi_2) = \widehat{F_1}(\xi_1, \xi_2)\mathbb{1}_{[-R,R]}(\xi_1)$$

for each $(\xi_1, \xi_2) \in \mathbb{R}^2$. With B_δ defined by (1.1), split

$$B_\delta(F_1, F_2) = B_\delta(F_{1,R}, F_2) + B_\delta(G_{1,R}, F_2). \tag{1.4}$$

Using Theorem 1.1 we estimate

$$\|B_\delta(G_{1,R}, F_2)\|_{\mathrm{L}^1(\mathbb{R}^2)} \lesssim_\zeta R^{-\sigma} \|G_{1,R}\|_{\mathrm{L}^2(\mathbb{R}^2)} \|F_2\|_{\mathrm{L}^2(\mathbb{R}^2)} \leq R^{-\sigma} \|F_1\|_{\mathrm{L}^2(\mathbb{R}^2)} \|F_2\|_{\mathrm{L}^2(\mathbb{R}^2)} \tag{1.5}$$

with $\sigma > 0$. It remains to control $B_\delta(F_{1,R}, F_2)$. Applying the Cauchy–Schwarz inequality in (x, y) for each fixed t, we obtain

$$\|B_\delta(F_{1,R}, F_2)\|_{\mathrm{L}^1(\mathbb{R}^2)} \lesssim_\zeta \|F_{1,R}(x+\delta, y) - F_{1,R}(x, y)\|_{\mathrm{L}^2_{(x,y)}(\mathbb{R}^2)} \|F_2\|_{\mathrm{L}^2(\mathbb{R}^2)}.$$

The Plancherel identity gives

$$\|F_{1,R}(x+\delta, y) - F_{1,R}(x, y)\|^2_{\mathrm{L}^2_{(x,y)}(\mathbb{R}^2)} = \int_{[-R,R]\times\mathbb{R}} |\widehat{F_1}(\xi_1, \xi_2)|^2 |e^{2\pi i\delta\xi_1} - 1|^2 \, d\xi_1 \, d\xi_2,$$

while $|\xi_1| \leq R$ implies that $|e^{2\pi i\delta\xi_1} - 1| \lesssim \delta R$. Therefore,

$$\|B_\delta(F_{1,R}, F_2)\|_{\mathrm{L}^1(\mathbb{R}^2)} \lesssim_\zeta \delta R \, \|F_1\|_{\mathrm{L}^2(\mathbb{R}^2)} \|F_2\|_{\mathrm{L}^2(\mathbb{R}^2)}. \tag{1.6}$$

From (1.5), (1.6) and the splitting (1.4), we finally conclude that

$$\|B_\delta(F_1, F_2)\|_{\mathrm{L}^1(\mathbb{R}^2)} \lesssim_\zeta (\delta R + R^{-\sigma}) \|F_1\|_{\mathrm{L}^2(\mathbb{R}^2)} \|F_2\|_{\mathrm{L}^2(\mathbb{R}^2)},$$

so the proof is completed by choosing $R = \delta^{-1/2}$ and $\gamma = \min\{1/2, \sigma/2\}$. □

This completes the proof of Proposition 2.1.

Acknowledgements. The authors are grateful to Terence Tao for raising the question answered here, and for pointing out its connection with [12]. The first author was supported by National Science Foundation grant DMS-1901413. The third author was supported in part by the Croatian Science Foundation under the project IP-2018-01-7491 (DEPOMOS).

REFERENCES

[1] I. Assani. Pointwise convergence of ergodic averages along cubes. *J. Anal. Math.* **110** (2010), 241–269.

[2] T. Austin. Norm convergence of continuous-time polynomial multiple ergodic averages. *Ergod. Th. & Dynam.* **32**(2) (2012), 361–382.

[3] V. Bergelson and A. Leibman. A nilpotent Roth theorem. *Invent. Math.* **147**(2) (2002), 429–470.

https://doi.org/10.1017/etds.2021.45 Published online by Cambridge University Press

[4] V. Bergelson and A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.* **9**(3) (1996), 725–753.
[5] V. Bergelson, A. Leibman and C. G. Moreira. From discrete- to continuous-time ergodic theorems. *Ergod. Th. & Dynam.* **32**(2) (2012), 383–426.
[6] J. Bourgain. Double recurrence and almost sure convergence. *J. Reine Angew. Math.* **404** (1990), 140–161.
[7] J. Bourgain. On the pointwise ergodic theorem on L^p for arithmetic sets. *Israel J. Math.* **61**(1) (1988), 73–84.
[8] J. Bourgain. Pointwise ergodic theorems for arithmetic sets. *Publ. Math. Inst. Hautes Études Sci.* **69** (1989), 5–45, with an appendix by the authors H. Furstenberg, Y. Katznelson and D. S. Ornstein.
[9] Z. Buczolich and R. D. Mauldin. Divergent square averages. *Ann. of Math. (2)* **171**(3) (2010), 1479–1530.
[10] A.-P. Calderón. Ergodic theory and translation-invariant operators. *Proc. Natl Acad. Sci. USA* **59** (1968), 349–353.
[11] M. Christ. On trilinear oscillatory integral inequalities and related topics. *Preprint*, 2020, arXiv:2007.12753.
[12] M. Christ, P. Durcik and J. Roos. Trilinear smoothing inequalities and a variant of the triangular Hilbert transform. *Preprint*, 2020, arXiv:2008.10140.
[13] Q. Chu and N. Frantzikinakis. Pointwise convergence for cubic and polynomial multiple ergodic averages of non-commuting transformations. *Ergod. Th. & Dynam.* **32**(3) (2012), 877–897.
[14] S. Donoso and W. Sun. A pointwise cubic average for two commuting transformations. *Israel J. Math.* **216**(2) (2016), 657–678.
[15] S. Donoso and W. Sun. Pointwise convergence of some multiple ergodic averages. *Adv. Math.* **330** (2018), 946–996.
[16] S. Donoso and W. Sun. Pointwise multiple averages for systems with two commuting transformations. *Ergod. Th. & Dynam.* **38**(6) (2018), 2132–2157.
[17] N. Frantzikinakis. Joint ergodicity of sequences. *Preprint*, 2021, arXiv:2102.09967.
[18] N. Frantzikinakis. Some open problems on multiple ergodic averages. *Bull. Hellenic Math. Soc.* **60** (2016), 41–90.
[19] N. Frantzikinakis, E. Lesigne and M. Wierdl. Random sequences and pointwise convergence of multiple ergodic averages. *Indiana Univ. Math. J.* **61**(2) (2012), 585–617.
[20] W. Huang, S. Shao and X. Ye. Pointwise convergence of multiple ergodic averages and strictly ergodic models. *J. Anal. Math.* **139**(1) (2019), 265–305.
[21] R. L. Jones, A. Seeger and J. Wright. Strong variational and jump inequalities in harmonic analysis. *Trans. Amer. Math. Soc.* **360**(12) (2008), 6711–6742.
[22] B. Krause, M. Mirek and T. Tao. Pointwise ergodic theorems for non-conventional bilinear polynomial averages. *Preprint*, 2020, arXiv:2008.00857.
[23] P. LaVictoire. Universally L^1-bad arithmetic sequences. *J. Anal. Math.* **113** (2011), 241–263.
[24] R. O'Neil. Convolution operators and $L(p,q)$ spaces. *Duke Math. J.* **30** (1963), 129–142.
[25] V. S. Varadarajan. Groups of automorphisms of Borel spaces. *Trans. Amer. Math. Soc.* **109** (1963), 191–220.
[26] M. N. Walsh. Norm convergence of nilpotent ergodic averages. *Ann. of Math. (2)* **175**(3) (2012), 1667–1688.