New York Journal of Mathematics

New York J. Math. **28** (2022) 1448–1462.

On the convergence of multiple ergodic means

Grigori A. Karagulyan, Michael T. Lacey and Vahan A. Martirosyan

Abstract. Consider a sequence of measure preserving transformations $U = \{U_k : k = 1, 2, \ldots\}$ on a measurable space $(X,)$. We prove a.e. convergence of the ergodic means

$$\frac{1}{s_1 \cdots s_n}\sum_{j_1=0}^{s_1-1}\cdots\sum_{j_n=0}^{s_n-1} f\left(U_1^{j_1}\cdots U_n^{j_n}x\right) \tag{0.1}$$

as $\min_j s_j \to \infty$, for any function $f \in L\log^{d-1}(X)$, where $d \le n$ is the rank of the transformations U. The result gives a generalization of a theorem by N. Dunford and A. Zygmund, claiming the convergence of (0.1) in a narrower class of functions $L\log^{n-1}(X)$.

Contents

1. Introduction

Birkho's ergodic theorem is one of the most important and beautiful result of probability theory. The study of ergodic theorems started in 1931 by von Neumann and Birkho, having its origins in statistical mechanics. Recall the denition of the measure-preserving transformation (see [4]).

Received July 30, 2022.

2010 *Mathematics Subject Classication.* 37A30, 37A46, 42B25.

Key words and phrases. Ergodic theorems, strong maximal function, multiple ergodic sums.

Research of Lacey was supported in part by grant from the US National Science Foundation, DMS-1949206, and the research of Karagulyan was supported by the Science Committee of RA, in the frames of the research project 21AG-1A045.

ISSN 1076-9803/2022

Denition 1.1. Let $(X, B,)$ be a probability space. A mapping $T : X \to X$ is said to be a measure-preserving transformation if for any measurable set $E \in B$ the set $T^{-1}(E)$ is also measurable and $(E) = (T^{-1}(E))$. The combination $(X, B, , T)$ is called a measure-preserving system.

Theorem A (Birkho)**.** *If* $(X, B, , T)$ *is a measure-preserving system, then for any function* $f \in L^1(X)$ *the averages*

$$\frac{1}{n}\sum_{j=0}^{n-1} f(T^j x)$$

converge almost everywhere to a T*-invariant function* f^* *as* $n \to \infty$*.*

There are dierent proofs and various generalizations of this classical theorem. Some of those clearly demonstrate strong link between the Lebesgue dierentiation theory on R^n and pointwise convergence of dierent type of ergodic averages. The following multiple version of Birkho's theorem, proved by Zygmund [13] and Dunford [2] independently, is an example of such a resemblance. Let $\Phi : R^+ \to R^+$ be non-decreasing function and $(X, B,)$ be a probability space. Denote by $L_\Phi(X)$ the class of B-measurable functions f on X with $\Phi(|f|) \in L^1(T)$. The class $L_\Phi(X)$ corresponding to a function

$$\Phi(t) = t\,(1 + (\max\{0, \log t\})^n)\,, \quad n \geq 1, \tag{1.1}$$

will be denoted by $L \log^n L(X)$. Clearly this class of function is strongly included in $L^1(X)$.

Theorem B (Dunford-Zygmund)**.** *Let* $U_1, \ldots, U_n$ *be measure-preserving one-to-one transformations of a probability space* $(X, B,)$*. Then for any function* $f \in L \log^{n-1} L(X)$ *the averages*

$$\frac{1}{s_1 \cdots s_n}\sum_{j_1=0}^{s_1-1} \cdots \sum_{j_n=0}^{s_n-1} f\left(U_1^{j_1} \cdots U_n^{j_n} x\right) \tag{1.2}$$

converge a.e. as $\min_j s_j \to \infty$*.*

This result has been generalized for general contraction operators on L^1, considering those instead of the operators $f \to f \circ U_k$ generated by the measure-preserving transformations U_k (Dunford-Schwartz [3], Fava [5]). Hagelstein and Stokolos in [10] proved the sharpness of the class of functions $L \log^{n-1} L(X)$ in the context of Theorem B. Namely,

Theorem C (Hagelstein-Stokolos)**.** *Suppose a collection of invertible commuting measure-preserving transformations* $U = \{U_k : k = 1, 2, \ldots, n\}$ *is non-periodic, that is for any non-trivial collection of integers* $p_k \in Z$*,* $k = 1, 2, \ldots, n$ *we have*

$$\{U_1^{p_1} \circ \ldots \circ U_n^{p_n}(x) = x\} = 0.$$

If $\Phi(t) = o(t \log^{n-1} t)$ *as* $t \to \infty$*, then there exists a function* $f \in L_\Phi(X)$ *such that averages* (1.2) *unboundedly diverge a.e..*

Denition 1.2. A set of invertible commuting measure-preserving (ICMP) transformations $U = \{U_k : k = 1, 2, \ldots, n\}$ is said to be *dependent* if there is a non-trivial collection of integers $p_k \in \mathbb{Z}$, $k = 1, 2, \ldots, n$, such that

$$(U_1^{p_1} \circ \ldots \circ U_n^{p_n})(x) = x \tag{1.3}$$

almost everywhere on X. If there is no such a collection of integers p_k, then we say U is *independent*. The *rank of* U denoted by $\operatorname{rank}(U)$ will be called the largest integer r for which there is an independent subset of cardinality r in U.

Remark 1.3. Note that according to our denition, the independence of U requires the failure of (1.3) on a set of positive measure for any non-trivial collection of integers $\{p_k\}$, while the condition of non-periodicity in Theorem C is a stronger version of independence, since in this case the failure of (1.3) is required almost everywhere.

The main result of the present paper provides a generalization of Theorem B. Namely, it says that in fact a.e. convergence of averages (1.2) holds in a larger class of functions $L \log^{d-1} L \supset L \log^{n-1} L$, where $d = \operatorname{rank}(U) \le n$. First we prove the following weak type maximal inequality, where $\mathrm{Log}_n t$ denotes the function in (1.1), i.e.

$$\mathrm{Log}_n(t) = t\,(1 + (\max\{0, \log t\})^n)\,.$$

Theorem 1.4. *Let* $U = \{U_k : k = 1, 2, \ldots, n\}$ *be a set of ICMP transformations of rank* d*. Then, for any function* $f \in L \log^{d-1} L(X)$ *and* $\lambda > 0$*, we have*

$$\begin{aligned}\left|\left\{x \in X : \sup_{s_j \ge 0} \frac{1}{s_1 \ldots s_n} \sum_{k_1=0}^{s_1-1} \cdots \sum_{k_n=0}^{s_n-1} f\left((U_1^{k_1} \circ \cdots \circ U_n^{k_n})(x)\right) > \lambda\right\}\right| \\ \le C(U) \int_X \mathrm{Log}_{d-1}\left(\frac{|f|}{\lambda}\right),\end{aligned} \tag{1.4}$$

where $C(U)$ *is a constant depending only on* U*.*

As a corollary of (1.4) we obtain the following.

Theorem 1.5. *Let* $U = \{U_k : k = 1, 2, \ldots, n\}$ *be a set of ICMP transformations of rank* d*. Then, for any function* $f \in L \log^{d-1} L(X)$ *the averages* (1.2) *converge almost everywhere as* $\min s_k \to \infty$*.*

Remark 1.6. We will see in the last section that the class $L \log^{d-1} L(X)$ of the functions in Theorem 1.5 is optimal. More precisely, if the corresponding independent subset of cardinality $d = \operatorname{rank}(U)$ in U is "strongly independent" (i.e. non-periodic), then under the condition $\Phi(t) = o(t \log^{d-1} t)$ there exists a function $f \in L_\Phi(X)$ with a.e. diverging averages (1.2). In fact, the proof of this optimality immediately follows from Theorem C. We will just need to apply a simple lemma proved in Section 5 (Lemma 5.1).

The inequality (1.4) will be deduced from a maximal inequality on $\mathbb{R}^n$. Let $A : \mathbb{R}^n \to \mathbb{R}^d$ be a linear operator given by the matrix

$$A = \{a_{kj} : 1 \le j \le n, 1 \le k \le d\} \tag{1.5}$$

of size $d \times n$ (d-rows and n-columns). We consider the maximal function

$$M_A f(\mathbf{x}) = \sup_R \frac{1}{|R|} \int_R |f(\mathbf{x} + A \cdot \mathbf{t})| d\mathbf{t}, \quad \mathbf{x} \in \mathbb{R}^d, \tag{1.6}$$

where sup is taken over all n-dimensional symmetric intervals

$$R = \{\mathbf{t} = (t_1, \dots, t_n) \in \mathbb{R}^n : t_j \in [-r_j, r_j],\ j = 1, 2, \dots, n\} \subset \mathbb{R}^n.$$

Denote by $\operatorname{rank} A$ the rank of the matrix A.

Theorem 1.7. *Let* A *be the matrix* (1.5) *and* $r = \operatorname{rank} A$. *Then for any function* $f \in L(\log^+ L)^{r-1}(\mathbb{R}^d)$ *the bound*

$$|\{\boldsymbol{x} \in \mathbb{R}^d : M_A f(\boldsymbol{x}) > \lambda\}| \le C(A) \int_{\mathbb{R}^d} \mathrm{Log}_{r-1}\left(\frac{|f|}{\lambda}\right), \tag{1.7}$$

holds, where $C(A)$ *is a constant, depending only the matrix* A.

Remark 1.8. Observe that if $n = d = r$ and A is the identity matrix of size n, then (1.6) gives the well-known strong maximal function on $\mathbb{R}^n$, correspondingly, (1.7) becomes the weak type inequality due to M. de Guzman [6] (see also [7]). Moreover, inequality (1.7) holds even if A is a general invertible matrix and it follows from Guzman's inequality of [6], simply using the equivalence of rectangular and parallelepiped differentiation bases on $\mathbb{R}^n$. Our proof of the full version of inequality (1.7) is a reduction of the general case to the case of invertible A.

Remark 1.9. Note that papers [2] and [13] suggest different proofs of Theorem B. The proof of [2] is straightforward and the convergence of averages (1.2) was established only for the functions in L^p, $1 < p < \infty$, while Zygmund [13] provides an inequality, which is the analogue of a similar inequality for the strong maximal function, originally proved in [9]. The latter is the weaker version of Guzman's inequality of [6] .

Remark 1.10. The well known transfer principle of Calderón [1] enables to reduce certain ergodic maximal inequalities to maximal inequalities in harmonic analysis. A version of Calderón's principle in higher dimension was suggested in [11], where only non-periodic collections of measure-preserving transformations were considered. In fact, our proof of Theorem 1.4 is an extension of this higher dimensional principle to arbitrary collections of measure-preserving transformations.

The authors are grateful to the unknown referee for valuable remarks.

2. Proof of Theorem 1.7

We will use the following equivalent form of the maximal function (1.6)

$$M_U f(\mathbf{x}) = \sup_{r_k>0} \frac{1}{2^n r_1 \cdots r_n} \int_{-r_1}^{r_1} \cdots \int_{-r_n}^{r_n} |f(\mathbf{x}+t_1\mathbf{u}_1+\cdots+t_n\mathbf{u}_n)|\,dt_1\cdots dt_n, \quad (2.1)$$

where the vector set $U = \{\mathbf{u}_k,\ k = 1,2,\dots,n\}$ is formed by the columns of the matrix (1.5). So the rank of vectors U coincides with the rank of the matrix A. Once again note that that if the collection of vectors are independent, i.e. the matrix A is invertible, then inequality (1.7) is known, and we are going to reduce the general case to the case of invertible A. We need several lemmas, concerning parallelepipeds in $\mathbb{R}^d$ and associated measures.

For a vector $\mathbf{x} = (x_1,\dots,x_d) \in \mathbb{R}^d$ we denote $|\mathbf{x}| = (x_1^2+\dots+x_d^2)^{1/2}$. Given a set of vectors $V \subset \mathbb{R}^d$ we denote by $\operatorname{span}(V)$ the linear space generated by V (sometimes this Euclidean space will be denoted by $\mathbb{R}_V$). The notation $|E|$ will stand for the Lebesgue measure of a set E in an Euclidean space.

Definition 2.1. Let $U = \{\mathbf{u}_k:\ k = 1,2,\dots,n\} \subset \mathbb{R}^d$ be a set of unit vectors. Call a parallelepiped in $\mathbb{R}^d$ a set of the form

$$R = \left\{\mathbf{x} \in \mathbb{R}^d :\ \mathbf{x} = t_1\mathbf{u}_1+\dots+t_n\mathbf{u}_n,\ t_j \in [-r_j, r_j]\right\}. \quad (2.2)$$

The family of all parallelepipeds (2.2) generated by a fixed set of vectors U will be denoted by $\mathcal{P}_U$.

Note that parallepipeds can have different representations (2.2). Clearly the arithmetic sum of two parallelepipeds R, Q

$$R+Q = \{\mathbf{x}+\mathbf{t}:\ \mathbf{x} \in R,\ t \in Q\}$$

is again a parallelepiped. For two parallelepipeds R and Q we write $Q \succ R$ if there is a parallelepiped R' such that $Q = R+R'$.

Lemma 2.2. *If* $U = \{\mathbf{u}_k:\ k = 1,2,\dots,n\}$ *is a basis set of vectors in* $\mathbb{R}^n$ *and* $R \in \mathcal{P}_U$ *has a representation* (2.2)*, then*

$$\{\mathbf{x} \in \mathbb{R}^n :\ |\mathbf{x}| \le 1\} \subset \frac{C(U)}{\min_j r_j} R, \quad (2.3)$$

where $C(U)$ *is a constant, depending only on the set of vectors* U.

Proof. For any $j = 1,2,\dots,n$ we consider hyperplanes Γ_j^+ and Γ_j^- in $\mathbb{R}^n$ defined

$$\Gamma_j^{\pm} = \{\mathbf{x} = t_1\mathbf{u}_1+\dots+t_n\mathbf{u}_n :\ t_j = \pm r_j,\ t_i \in \mathbb{R},\ i \ne j\}$$

and let S_j be the closed strip domain lying between the hyperplanes $\Gamma_j^{\pm}$. We have $R = \cap_j S_j$. Denote by h_j the distance of the hyperplanes Γ_j^+ and Γ_j^- from the origin. It is clear that

$$\{\mathbf{x} \in \mathbb{R}^n :\ |\mathbf{x}| \le \min_j h_j\} \subset R. \quad (2.4)$$

One can also check that $c_j = h_j/r_j$ are constants, depending only on U. Denote $C(U) = (\min_j c_j)^{-1}$. From (2.4) we obtain

$$\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \le 1\} \subset \frac{1}{\min_j h_j} R \subset \frac{C(U)}{\min_j r_j} R$$

and so (2.3).

A version of the following lemma in the case of $d = 2$ was proved by Guzmán-Welland in [8] (see also [7], chap. 6, Lemma 2.1).

Lemma 2.3 (Guzmán-Welland). *Let* $U = \{\boldsymbol{u}_k : k = 1, 2, \dots, n\}$ *be a set of unit vectors in* $\mathbb{R}^d$. *Then for any parallelepiped* $R \in \mathcal{P}_U$ *there exist a subset* $V \subset U$ *of independent vectors and a parallelepiped* $Q \in \mathcal{P}_V$ *such that*

$$\operatorname{rank}(V) = \operatorname{rank}(U), \tag{2.5}$$

$$Q \subset R, \tag{2.6}$$

$$R \subset C(U)\, Q, \tag{2.7}$$

where $C(U)$ *is a constant depending only on the set of vectors* U.

Proof. Suppose that $R \in \mathcal{P}_U$ is the parallelepiped (2.2). Without loss of generality we can suppose that

$$r_1 \ge r_2 \ge \dots \ge r_n. \tag{2.8}$$

Denote

$$V = \{\mathbf{u}_k : \mathbf{u}_k \notin \operatorname{span}\{\mathbf{u}_1, \dots, \mathbf{u}_{k-1}\}\} \subset U.$$

One can easily check that the vectors of V are independent and $\operatorname{rank}(V) = \operatorname{rank}(U)$. One can split the set of vectors U into groups

$$U_j = \{\mathbf{u}_k : k \in (k_{j-1}, k_j]\}, \quad j = 1, 2, \dots, s,$$
$$0 = k_0 < k_1 < \dots < k_s = n,$$

such that

$$V = \bigcup_{i \ge 0} U_{2i+1}, \quad U_{2j} \subset \operatorname{span}\left(\bigcup_{i=1}^{j} U_{2i-1}\right).$$

Considering the parallelepipeds

$$R_j = \left\{\mathbf{x} \in \mathbb{R}^d : \mathbf{x} = \sum_{k=k_{j-1}+1}^{k_j} t_k \mathbf{u}_k,\ t_k \in [-r_k, r_k]\right\} \in \mathcal{P}_{U_j},$$

we can write

$$R = R_1 + R_2 + \dots + R_s.$$

Then the parallelepiped

$$Q = \sum_{j:\, 2j-1 \le s} R_{2j-1}$$

satises (2.5) and (2.6). If $\mathbf{x} \in R_{2j}$, then

$$|\mathbf{x}| \le \sum_{i=k_{2j-1}+1}^{k_{2j}} r_i \le n r_{k_{2j-1}}. \tag{2.9}$$

Let Y_j be the subspace of $\mathbb{R}^d$ generated by the independent vectors $\bigcup_{i\le j} U_{2i-1}$. One can check

$$R_i \subset Y_j, \quad i = 1, 2, \dots, 2j.$$

Thus, applying Lemma 2.2 for the space Y_j, as well as (2.8), (2.9), we conclude

$$\frac{1}{n r_{k_{2j-1}}} R_{2j} \subset \{\mathbf{x} \in Y_j : |\mathbf{x}| \le 1\} \subset \frac{C(U)}{r_{k_{2j-1}}}(R_1 + R_3 + \dots + R_{2j-1}) \subset \frac{C(U)}{r_{k_{2j-1}}} Q$$

Thus we get $R_{2j} \subset nC(U) Q$ and therefore

$$R \subset n^2 C(U) Q.$$

This gives us (2.7), completing the proof of lemma.

Given a set of unit vectors $U = \{\mathbf{u}_k : k = 1, 2, \dots, n\} \subset \mathbb{R}^d$, let $\mathbb{R}_U$ be the sub-space of $\mathbb{R}^d$ generated by the vectors U. We associate with a parallelepiped (2.2) a probability measure μ_R supported on R as follows. First, for each j we con-sider a probability measure μ_j uniformly distributed on the one dimensional parallelepiped $\{t\mathbf{u}_j : t \in [-r_j, r_j]\}$. The convolution of singular measures μ_j is the measure μ_R dened on the Lebesgue measurable sets of $E \subset \mathbb{R}_U$ by

$$\mu_R(E) = \int_{\mathbb{R}_U} \dots \int_{\mathbb{R}_U} \mathbf{1}_E(\mathbf{v}_1 + \dots + \mathbf{v}_n) d\mu_1(\mathbf{v}_1) \dots d\mu_n(\mathbf{v}_n). \tag{2.10}$$

One can check that μ_R is well-dened for any Lebesgue measurable set $E \subset \mathbb{R}_U$. Denote by f_R the density function of measure μ_R with respect to the Lebesgue measure on $\mathbb{R}_U$. Observe that if U is independent, then

$$f_R(\mathbf{x}) = \begin{cases} |R|^{-1} & \text{if } \mathbf{x} \in R, \\ 0 & \text{if } \mathbf{x} \in \mathbb{R}_U \setminus R. \end{cases} \tag{2.11}$$

Lemma 2.4. *Let* $U \subset \mathbb{R}^d$ *be a set of arbitrary unit vectors and* $R \in \mathcal{P}_U$. *Then there exists a set of independent vectors* $V \subset U$ *such that* $\mathrm{rank}(V) = \mathrm{rank}(U)$ *and there is a parallelepiped* $R^* \in \mathcal{P}_V$ *such that*

$$\mu_R \le C(U)\, \mu_{R^*}. \tag{2.12}$$

Proof. Applying Lemma 2.3 in the Euclidean space $\mathbb{R}_U$, we nd a set of in-dependent vectors $V \subset U$, $\mathrm{rank}(V) = \mathrm{rank}(U)$ and a parallelepiped $Q \in \mathcal{P}_V$ satisfying the conditions of lemma. Since $Q \subset R$, we have $R = Q + H$ for some parallelepiped H in $\mathbb{R}_U$. We can write

$$\mu_R(E) = \int_{\mathbb{R}_U} \int_{\mathbb{R}_U} \mathbf{1}_E(\mathbf{v} + \mathbf{v}^*) d\mu_Q(\mathbf{v}) d\mu_H(\mathbf{v}^*)$$

$$= \int_{R_U}\int_{R_U} \mathbf{1}_E(\mathbf{v}+\mathbf{v}')f_Q(\mathbf{v})\,d\mathbf{v}\,d\mu_H(\mathbf{v}')$$
$$= \frac{1}{|Q|}\int_{R_U}\int_{R_U} \mathbf{1}_E(\mathbf{v}+\mathbf{v}')\mathbf{1}_Q(\mathbf{v})\,d\mathbf{v}\,d\mu_H(\mathbf{v}')$$
$$\le \frac{|E|}{|Q|}.$$

This clearly implies

$$\|f_R\|_\infty \le \|f_Q\|_\infty = |Q|^{-1}. \tag{2.13}$$

Denote $R' = C(U)Q$, where $C(U)$ is the constant in (2.7). From (2.7) and (2.11) we have

$$R \subset R',$$
$$\|f_{R'}\|_\infty = |R'|^{-1} = (C(U)|Q|)^{-1}. \tag{2.14}$$

Combining (2.13) and (2.14) we get the pointwise bound $f_R \le C(U)f_{R'}$, which implies (2.12).

Proof of Theorem 1.7. Observe that the integral in (2.1) may be written as a convolution of measure (2.10) with the function f. Namely, we have

$$\frac{1}{2^n r_1\dots r_n}\int_{-r_1}^{r_1}\dots\int_{-r_n}^{r_n} |f(\mathbf{x}+t_1\mathbf{u}_1+\dots+t_n\mathbf{u}_n)|\,dt_1\dots dt_n$$
$$= \int_{\mathbb{R}^d} |f(\mathbf{x}+\mathbf{v})|\,d\mu_R(\mathbf{v}). \tag{2.15}$$

Applying Lemma 2.4, for any parallelepiped $R \in P_U$ we find an independent vector set $V \subset U$ with $\mathrm{rank}(V) = \mathrm{rank}(U)$ and a parallelepiped $R' \in P_V$ such that (2.12) holds. Thus the last integral in (2.15) may be estimated as follows:

$$\int_{\mathbb{R}^d} f(\mathbf{x}+\mathbf{v})\,d\mu_R(\mathbf{v}) \le C(U)\int_{\mathbb{R}^d} f(\mathbf{x}+\mathbf{v})\,d\mu_{R'}(\mathbf{v}) \le C(U)M_V f(\mathbf{x}).$$

This implies

$$M_U f(\mathbf{x}) \le C(U)\max_V M_V f(\mathbf{x}),$$

where the maximum is taken over all the subsets $V \subset U$ of independent vectors such that $\mathrm{rank}(V) = \mathrm{rank}(U)$. For each such V the operator M_V satisfies the bound (1.7) and the number of all collections V is constant, depending only on n and so on U. Thus we get (1.7).

3. A discrete maximal inequality

We will need a discrete version of inequality (1.7). Let $\varphi : \mathbb{Z}^d \to \mathbb{R}$ be a d-dimensional sequence and let $A = \{a_{kj} : 1 \le j \le n,\ 1 \le k \le d\}$ be an integer matrix. Consider the maximal operator

$$
\begin{aligned}
D_A\varphi(\mathbf{n}) &= \sup_{s_j\in\mathbb{N}} \frac{1}{s_1\dots s_n}\sum_{k_1=0}^{s_1-1}\dots\sum_{k_n=0}^{s_n-1}\varphi(\mathbf{n}+A\mathbf{k})\\
&= \sup_{s_j\in\mathbb{N}} \frac{1}{s_1\dots s_n}\sum_{\mathbf{k}=0}^{\mathbf{s}-1}\varphi(\mathbf{n}+A\mathbf{k}), \qquad \mathbf{n}\in\mathbb{N}^d.
\end{aligned}
$$

From Theorem 1.7 we easily obtain the following.

Corollary 3.1. *For any integer matrix* A *of* $\operatorname{rank}(A)=r$ *we have the bound*

$$
\#\{\boldsymbol{n}\in\mathbb{Z}^d:\ D_A\varphi(\boldsymbol{n})>\lambda\}\le C(A)\sum_{\boldsymbol{n}\in\mathbb{Z}^d}\frac{|\varphi(\boldsymbol{n})|}{\lambda}\operatorname{Log}_{r-1}\left(\frac{|\varphi(\boldsymbol{n})|}{\lambda}\right).
$$

Proof. Given multiple sequence $\varphi(\mathbf{m})$ consider the function

$$
f(\mathbf{x})=\sum_{\varepsilon_j=0,1,-1}\varphi(m_1+\varepsilon_1,\dots,m_n+\varepsilon_n),\ \text{if } [\mathbf{x}]=\mathbf{m},\qquad \mathbf{m}\in\mathbb{Z}_d, \tag{3.1}
$$

on $\mathbb{R}^d$, where $[\mathbf{x}]=([x_1],\dots,[x_d])$ denotes the coordinate wise integer part of the vector $\mathbf{x}=(x_1,\dots,x_d)$. Clearly there is a constant $\delta=\delta(A)<1$ such that

$$
A(\Delta)\subset(-1,1)^d,\ \text{where } \Delta=[0,\delta)^n, \tag{3.2}
$$

Using (3.1), (3.2), one can check that

$$
\varphi(\mathbf{n}+A\mathbf{k})\le f(\mathbf{x}+A\mathbf{t})\ \text{if } \mathbf{t}\in\mathbf{k}+\Delta,\ [\mathbf{x}]=\mathbf{n}.
$$

Thus we obtain

$$
\begin{aligned}
\sum_{\mathbf{k}=0}^{\mathbf{s}-1}\varphi(\mathbf{n}+A\mathbf{k}) &\le \sum_{\mathbf{k}=0}^{\mathbf{s}-1}\frac{1}{|\Delta|}\int_{\mathbf{k}+\Delta}|f(\mathbf{x}+A\mathbf{t})|d\mathbf{t}\\
&\le \frac{1}{|\Delta|}\int_R|f(\mathbf{x}+A\mathbf{t})|d\mathbf{t},
\end{aligned}
$$

for any $\mathbf{x}$ with $[\mathbf{x}]=\mathbf{n}$, where

$$
R=\{\mathbf{t}\in\mathbb{R}^n:\ t_j\in[-1,s_j],\ j=1,\dots,n\}.
$$

This implies

$$
D_A\varphi(\mathbf{n})\le C(A)M_Af(\mathbf{x})\ \text{if } [\mathbf{x}]=\mathbf{n}\in\mathbb{Z}^d
$$

and so

$$
\begin{aligned}
\#\{\mathbf{n}\in\mathbb{Z}_+^d:\ D_A\varphi(\mathbf{n})>\lambda\} &\le |\{\mathbf{x}\in\mathbb{R}^d:\ M_Af(\mathbf{x})>\lambda/C(A)\}|\\
&\le C(A)\int_{\mathbb{R}^d}\frac{|f|}{\lambda}\operatorname{Log}_{r-1}\left(\frac{|f|}{\lambda}\right)\\
&\le C(A)\sum_{\mathbf{n}\in\mathbb{Z}^d}\frac{|\varphi(\mathbf{n})|}{\lambda}\operatorname{Log}_{r-1}\left(\frac{|\varphi(\mathbf{n})|}{\lambda}\right).
\end{aligned}
$$

This completes the proof.

4. Proofs of Theorems 1.4 and 1.5

Proof of 1.4. Since $\mathrm{rank}(U) = d$, without loss of generality we can suppose that $U_1, \dots, U_d$ are independent and

$$U_k^{l_k} = U_1^{a_{1,k}} \circ \cdots \circ U_d^{a_{d,k}}, \quad d < k \le n, \tag{4.1}$$

where $l_k \ge 1$ and $a_{j,k}$ are some integers. First we suppose that $l_k = 1$. Thus we can write

$$\begin{aligned} f\,(U_1^{k_1} \circ \cdots \circ U_n^{k_n})(x) &= f\,(U_1^{k_1+a_{1,d+1}k_{d+1}+\cdots+a_{1,n}k_n} \circ \cdots \circ U_d^{k_d+a_{d,d+1}k_{d+1}+\cdots+a_{d,n}k_n})(x) \\ &= (x, A\,\mathbf{k}), \end{aligned} \tag{4.2}$$

where

$$(x, \mathbf{n}) = f\,(U_1^{n_1} \circ \cdots \circ U_d^{n_d})(x), \ \mathbf{n} = (n_1, \dots, n_d) \in \mathbb{Z}^d,$$

and

$$A = \begin{pmatrix} 1 & 0 & \cdots & 0 & a_{1,d+1} & \cdots & a_{1,n} \\ 0 & 1 & \cdots & 0 & a_{2,d+1} & \cdots & a_{2,n} \\ & & \cdots & & & \cdots & \\ 0 & 0 & \cdots & 1 & a_{d,d+1} & \cdots & a_{d,n} \end{pmatrix}$$

is a matrix of size $d \times n$. Let

$$f_M^*(x, \mathbf{n}) = \max_{1 \le s_j \le M} \frac{1}{s_1 \cdots s_n} \sum_{\mathbf{k}=0}^{\mathbf{s}-1} |(x, \mathbf{n} + A\,\mathbf{k})|, \tag{4.3}$$

where $M \in \mathbb{N}$ and denote

$$\begin{aligned} E(x) &= \{\mathbf{n} : 1 \le n_j \le N : f_M^*(x, \mathbf{n}) > \ \}, \\ E(\mathbf{n}) &= \{x : f_M^*(x, \mathbf{n}) > \ \}, \qquad \mathbf{n} \in \mathbb{Z}^d, \\ E &= \{(x, \mathbf{n}) : 1 \le n_j \le N, f_M^*(x, \mathbf{n}) > \ \} = \cup_{x \in X} E(x) \\ &= \cup_{1 \le n_j \le N} E(\mathbf{n}). \end{aligned} \tag{4.4}$$

Taking into account (4.2), observe that inequality (1.4) is the same as

$$\lim_{M \to \infty} (E(\mathbf{0})) \le C(U) \cdot \int_X \mathrm{Log}_{d-1} \left(\frac{|f|}{\ } \right). \tag{4.5}$$

In (4.3) the coordinates of $A\,\mathbf{k}$ may vary in the interval $[-R, R]$, where $R = R(A, M)$ is a constant depending only on the matrix A and the integer M. From Corollary 3.1 it follows that

$$\#(E(x)) \le C(A) \sum_{1 \le n_j \le N+R} \mathrm{Log}_{r-1} \left(\frac{|(x, \mathbf{n})|}{\ } \right) \text{ for all } x \in X.$$

Then, since U_k are measure-preserving, the sets $E(\mathbf{n})$ have equal measures for different $\mathbf{n} \in \mathbb{Z}^d$. Thus from (4.4) we obtain

$$
\begin{aligned}
(E(\mathbf{0})) &= \frac{1}{N^d}\sum_{1\le n_j\le N}(E(\mathbf{n})) = \frac{1}{N^d}\int_X \#(E(x)) \\
&\le \frac{C(A)}{N^d}\sum_{1\le n_j\le N+R}\int_X \mathrm{Log}_{r-1}\left(\frac{|(x,\mathbf{n})|}{}\right) \\
&= \frac{C(A)(N+R)^d}{N^d}\int_X \mathrm{Log}_{r-1}\left(\frac{|f|}{}\right).
\end{aligned}
$$

Fixing M and letting $N \to \infty$, we get

$$|E(\mathbf{0})| \le C(A)\int_X \mathrm{Log}_{r-1}\left(\frac{|f|}{}\right),$$

which implies (4.5). The general case $l_k \ge 1$ can be easily deduced from the case of $l_k = 1$. Fix an integer vector $\mathbf{r} = (r_{d+1},\dots,r_n)$, $0 \le r_j < l_j$, and denote by $Q^{\mathbf{r}}_{s_1,\dots,s_d} f(x)$ the sum of functions

$$f\left(U_1^{k_1}\cdots U_n^{k_n}x\right),$$

over the integer vectors $\mathbf{k} = (k_1,\dots,k_n)$, satisfying

$$1 \le k_j < s_j, \quad 1 < j \le n, \tag{4.6}$$

$$k_j = \tilde{k}_j l_j + r_j, \quad \tilde{k}_j \in \mathbb{Z}, \quad d < j \le n. \tag{4.7}$$

Under the conditions (4.7) we can write

$$f\left(U_1^{k_1}\cdots U_n^{k_n}x\right) = \tilde{f}\left(U_1^{k_1}\cdots U_d^{k_d}\tilde{U}_{d+1}^{\tilde{k}_{d+1}}\cdots \tilde{U}_n^{\tilde{k}_n}x\right) \tag{4.8}$$

where

$$\tilde{f}(x) = f\left(U_{d+1}^{r_{d+1}}\cdots U_n^{r_n}x\right),$$

$$\tilde{U}_j = U_j^{l_j}, \quad d < j \le n.$$

From (4.1) it follows that

$$\tilde{U}_k = U_1^{a_{1,k}}\circ\cdots\circ U_d^{a_{d,k}}, \quad d < k \le n, \tag{4.9}$$

Denote by $(\mathbf{s},\mathbf{r})$ the number of integer vectors $\mathbf{k} = (k_1,\dots,k_n)$, satisfying (4.6) and (4.7). According to (4.8) and (4.9) we can say that

$$\frac{Q^{\mathbf{r}}_{s_1,\dots,s_n} f(x)}{(\mathbf{s},\mathbf{r})} \tag{4.10}$$

are certain ergodic averages, obeying the case of $l_k = 1$ in (4.1). Thus we conclude that the averages (4.10) satisfy the weak estimate (1.4) for all vectors $\mathbf{r}$.

On the other hand, taking into account $|(\mathbf{s},\mathbf{r})| \le s_1 \dots s_n$, we have

$$
\begin{aligned}
\frac{1}{s_1 \dots s_n}\sum_{k_1=0}^{s_1-1}\dots\sum_{k_n=0}^{s_n-1}\left| f\,(U_1^{k_1}\circ\dots\circ U_n^{k_n})(x)\right|
&= \frac{1}{s_1\dots s_n}\sum_{\mathbf{r}} Q^{\mathbf{r}}_{s_1,\dots,s_n} f(x)\\
&= \sum_{\mathbf{r}} \frac{|(\mathbf{s},\mathbf{r})|}{s_1\dots s_n}\,\frac{Q^{\mathbf{r}}_{s_1,\dots,s_n} f(x)}{|(\mathbf{s},\mathbf{r})|}\\
&\le \sum_{\mathbf{r}} \frac{Q^{\mathbf{r}}_{s_1,\dots,s_n} f(x)}{|(\mathbf{s},\mathbf{r})|}.
\end{aligned}
$$

Thus, since the averages (4.10) satisfy the weak estimate (1.4) and the number of dierent vectors $\mathbf{r} = (l_{d+1}, \dots, l_n)$ is a constant depending on $\mathcal{U}$ only, we obtain (1.4) in full generality. The theorem is proved.

Proof of Theorem 1.5. According to Theorem B the averages (1.2) converge a.e. for any function from $L\log^{n-1}L$ and so for any $f \in L^{\infty}(X)$. To prove convergence for any $f \in L\log^{d-1}L(T)$, fix $\varepsilon > 0$ and choose a function $g \in L^{\infty}$ such that

$$
\int_X \mathrm{Log}_{d-1}\left(\frac{|f-g|}{\varepsilon}\right) < \varepsilon.
$$

Applying (1.4), for the averages

$$
A_{\mathbf{m}}(f) = \frac{1}{m_1\dots m_n}\sum_{j_1=0}^{m_1-1}\dots\sum_{j_n=0}^{m_n-1} f\left(U_1^{j_1}\dots U_n^{j_n}x\right)
$$

we obtain

$$
\begin{aligned}
&\left\{x : \limsup_{\min n_j\to\infty} |A_{\mathbf{n}}(f) - A_{\mathbf{m}}(f)| > 2\varepsilon\right\}\\
&= \left\{x : \limsup_{\min n_j\to\infty} |A_{\mathbf{n}}(f-g) - A_{\mathbf{m}}(f-g)| > 2\varepsilon\right\}\\
&\le \left\{x : \sup_{\mathbf{n}} |A_{\mathbf{n}}(f-g)| > \varepsilon\right\}\\
&\le C(\mathcal{U})\int_X \mathrm{Log}_{d-1}\left(\frac{|f-g|}{\varepsilon}\right) < C(\mathcal{U})\varepsilon.
\end{aligned}
$$

This implies a.e. convergence of $A_{\mathbf{n}}(f)$, completing the proof of the theorem.

5. Sharpness in Theorem 1.5 and an extension

Let us show that the class of functions $L\log^{d-1}L(X)$ in Theorem 1.5 is optimal. Suppose the rank of $U = \{U_k : k = 1, 2, \dots, n\}$ is d and $\{U_1, \dots, U_d\}$ is the corresponding independent subset U, which is moreover non-periodic. According to Theorem C for $\Phi(t) = o(t\log^{d-1} t)$ there exists a function $f \in L_\Phi(X)$ with a.e. diverging averages

$$\frac{1}{s_1 \dots s_d}\sum_{j_1=0}^{s_1-1}\dots\sum_{j_d=0}^{s_d-1} f\left(U_1^{j_1}\dots U_d^{j_d}x\right). \tag{5.1}$$

It turns out that for the same function f we have a.e. divergence of the averages

$$\frac{1}{s_1 \dots s_n}\sum_{j_1=0}^{s_1-1}\dots\sum_{j_n=0}^{s_n-1} f\left(U_1^{j_1}\dots U_n^{j_n}x\right). \tag{5.2}$$

This immediately follows from the following lemma.

Lemma 5.1. *Let* $U = \{U_k : k = 1, 2, \dots, n\}$ *be a set of measure-preserving transformations and* $d \le n$. *If averages* (5.1) *diverge unboundedly a.e, then extended averages* (5.2) *also diverge unboundedly a.e.*

Proof. Denote by $A_{\mathbf{s}}(f)$ and $\tilde A_{\tilde{\mathbf{s}}}(f)$ the averages (5.1) and (5.2) respectively and consider the functions

$$M_p(f) = \max_{\mathbf{s}\in\mathbb{Z}_+^d,\, s_j\ge p} A_{\mathbf{s}}(f), \quad \tilde M_p(f) = \max_{\tilde{\mathbf{s}}\in\mathbb{Z}_+^n,\, s_j\ge p} \tilde A_{\tilde{\mathbf{s}}}(f).$$

The unbounded divergence of averages (5.1) implies $M_p(f) = \infty$ a.e. for any $p > 0$. If $\mathbf{s} = (s_1, \dots, s_d)$ and $\tilde{\mathbf{s}} = (s_1, \dots, s_d, \dots, s_n)$, then we have

$$\tilde A_{\tilde{\mathbf{s}}}(f) \ge \frac{A_{\mathbf{s}}(f)}{s_{d+1}\dots s_n},$$

and thus, for any $p > 0$

$$\tilde M_p(f) \ge \frac{1}{p^{n-d}} M_p(f) = \infty \text{ a.e.}.$$

A set of real numbers

$$\Theta = \{\theta_1, \theta_2, \dots, \theta_n\} \tag{5.3}$$

is said to be *dependent* (with respect to the rational numbers) if there is a nontrivial collection of integers r_k, $k = 1, 2, \dots, n$, such that

$$r_1\theta_1 + r_2\theta_2 + \dots + r_n\theta_n = 0 \mod 1,$$

If there are no such integers, then we say that Θ is *independent*. The rank of a collection $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ will be called the largest integer d, for which there is an independent subset of cardinality d in U. Consider the probability space of Lebesgue measure on $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ with modulo one addition. Applying Theorem

1.7 and the ergodicity of the rotation mapping x ™ x + for an irrational , we obtain

Corollary 5.2. *If* (5.3) *is a sequence of rank* d*, then*

1) for any f ¸ $L \log^{d*1} L(T)$ *the limit below holds a.e.*

$$\lim_{\min\{s_k\}™Ø} \frac{1}{s_1 5 s_n} \sum_{k_1=0}^{s_1*1} 5 \sum_{k_n=0}^{s_n*1} f(x + k_1 {}_1 + 5 + k_n {}_n) = \int_T f(x)dx, \tag{5.4}$$

2) for any increasing function $\Phi : R^+ ™ R^+$*, satisfying* $\Phi(t) = o(t(\log t)^{d*1})$*, there exists a function* f ¸ $L_\Phi(T)$ *such that the averages in* (5.4) *are a.e. divergent as* $\min\{s_k\}$ ™ Ø*.*

References

[1] Calderón, A.-P. Ergodic theory and translation-invariant operators. *Proc. Nat. Acad. Sci. U.S.A.* **59** (1968), 349–353. MR227354 (37 #2939), Zbl 0185.21806. 1451

[2] Dunford, Nelson. An individual ergodic theorem for non-commutative transformations, I. *Acta Sci. Math. (Szeged)* **14** (1951), 1–4. MR42074 (13,49f) , Zbl 0044.12501. 1449, 1451

[3] Dunford, Nelson; Schwartz, Jacob T. Linear operators. Part I.*John Wiley & Sons, Inc., New York*, 1988. MR1009164 (90g:47001a). 1449

[4] Einsiedler, Manfred; Ward, Thomas. Ergodic theory with a view towards number theory. Graduate Texts in Mathematics, 259, *Springer-Verlag London, Ltd., London*, 1984. MR2723325 (2012d:37016). 1448

[5] Fava, Norberto Angel. Weak type inequalities for product operators. *Studia Math.* **42** (1972), 271–288. MR308364 (46 #7478), Zbl 0237.47006. 1449

[6] de GuzmÆn, Miguel. An inequality for the Hardy-Littlewood maximal operator with respect to a product of dierentiation bases. *Studia Math.* **49** (1973/74), 185–194. MR333093 (48 #11418), Zbl 0286.28003. 1451

[7] de GuzmÆn, Miguel. Dierentiation of integrals in R^n. *Springer-Verlag, Berlin-New York*, 1975. MR0457661 (56 #15866), Zbl 0327.26010. 1451, 1453

[8] de GuzmÆn, Miguel; Welland, Grant V. On the dierentiation of integrals. *Rev. Un. Mat. Argentina* **25** (1970/71), 253–276. MR318418 (47 #6965), Zbl 0325.28004. 1453

[9] Jessen, B.; Marcinkiewicz, J.; Zygmund, A. Note on the dierentiability of multiple integrals. *Fund. Math.* **25** (1935), 235–252. 1451

[10] Hagelstein, Paul; Stokolos, Alexander. Weak type inequalities for ergodic strong maximal operators. *Acta Sci. Math. (Szeged)* **76** (2010), no. 3-4, 427–441. MR2789679 (2012c:47028). 1449

[11] Hagelstein, Paul; Stokolos, Alexander. Transference of weak type bounds of multiparameter ergodic and geometric maximal operators. *Fund. Math.* **218** (2012), no. 3, 269–284. MR2982778, Zbl 1257.37011. 1451

[12] Wiener, Norbert. The ergodic theorem. *Duke Math. J.* **5** (1939), 1–18. MR1546100, Zbl 0021.23501.

[13] Zygmund, A. An individual ergodic theorem for non-commutative transformations. *Acta Sci. Math. (Szeged)* **14** (1951), 103–110. MR45948, Zbl 0045.06403. 1449, 1451

(G. A. Karagulyan) Institute of Mathematics NAS RA, 24/5 Marshal Baghramian ave. Yerevan, 0019, Republic of Armenia, and Faculty of Mathematics and Mechanics, Yerevan State University, Alex Manoogian, 1, Yerevan, 0025, Republic of Armenia
g.karagulyan@ysu.am

(M. T. Lacey) School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA
lacey@math.gatech.edu

(V. A. Martirosyan) Faculty of Mathematics and Mechanics, Yerevan State University, Alex Manoogian, 1, 0025, Yerevan, Armenia
vahanmartirosyan2000@gmail.com

This paper is available via http://nyjm.albany.edu/j/2022/28-62.html.