



# On the speed of convergence in the ergodic theorem for shift operators

Nikolaos Chalmoukis<sup>ID</sup>, Leonardo Colzani, Bianca Gariboldi<sup>ID</sup>, and Alessandro Monguzzi<sup>ID</sup>

*Abstract.* Given a probability space  $(X, \mu)$ , a square integrable function  $f$  on such space and a (unilateral or bilateral) shift operator  $T$ , we prove under suitable assumptions that the ergodic means  $N^{-1} \sum_{n=0}^{N-1} T^n f$  converge pointwise almost everywhere to zero with a speed of convergence which, up to a small logarithmic transgression, is essentially of the order of  $N^{-1/2}$ . We also provide a few applications of our results, especially in the case of shifts associated with toral endomorphisms.

## 1 Introduction and main results

Let  $(X, \mu)$  be a probability space, and let  $T$  be a bounded linear operator on the Hilbert space  $L^2(X, \mu)$ . For  $f \in L^2(X, \mu)$ , consider its ergodic means

$$\frac{1}{N} \sum_{n=1}^{N-1} T^n f(x), \quad N \geq 1, x \in X.$$

In this article, we study the speed of convergence of such ergodic means when  $T$  is a unilateral or bilateral shift operator. Shift operators are sometimes induced by ergodic transformations. Thus, our results also cover some particular instances of von Neumann’s [vN32] and Birkhoff’s [Bir31] ergodic theorems. It is well-known that, in full generality, Birkhoff’s and von Neumann’s theorems are optimal, in the sense that the speed of convergence can indeed be arbitrarily slow, either in norm or in the sense of almost everywhere convergence (see [KP81, Kre79], cf. Theorem 1.2). Nonetheless, scholars have been intensively investigating such problems from different perspectives and with different goals in mind. To keep track of the literature, as it often happens, is a hard task and here we recall only a few meaningful papers, apologizing for the ones we omit. In [FS99], Furman and Shalom consider the measure-preserving and ergodic action of a locally compact group acting on a probability space and study the ergodic

---

Received by the editors December 14, 2023; revised May 8, 2024; accepted June 18, 2024.

All the authors are members of Indam–Gnampa. L.C., B.G., and A.M. are partially supported by the Indam–Gnampa project CUP\_E53C23001670001. N.C. and A.M. are partially supported by the Indam–Gnampa project CUP\_E53C22001930001 and by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “2nd Call for H.F.R.I. Research Projects to support Faculty Members & Researchers” (Project Number: 4662). A.M. and B.G. are supported by the PRIN 2022 project “TIGRECO – Time-varying signals on Graphs: REal and COmplex methods” funded by the European Union Next Generation – EU, Grant\_20227TRY8H, CUP\_F53D23002630001.

AMS subject classification: 47A35, 37A30.

Keywords: Ergodic theorem, convergence rate, shift operators, toral endomorphisms.



properties of the action along random walks on  $G$ . The setting described in [FS99] is quite different from ours, however, the results obtained are similar in the spirit with the ones we obtain here (cf. [FS99, Theorem 1.2] with Theorem 1.4). Kachurovskii, Podvigin, and coauthors have been studying the problem for the last decades from the spectral theory point of view and we refer the reader to the survey [KP16]. In the same spirit of the work of Kachurovskii and collaborators, we also mention the work [BAM21]. Avigad and collaborators investigated the rate of convergence in [AGT10, AI13, AR15] in the sense of *metastability* (see [Tao12]). Finally, we mention the work of Das and Yorke [DY18], of Bayart, Buczolich, and Heurieux [BBH20] and of Colzani, Gariboldi, and Monguzzi [Col22, CGM24], who all obtain the results on the speed of convergence when one considers as transformation the map  $x \rightarrow x + \alpha$ , which is an ergodic transformation of the  $d$ -dimensional torus  $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$  whenever  $\alpha = (\alpha_1, \dots, \alpha_d)$  is an irrational vector, that is, whenever  $1, \alpha_1, \dots, \alpha_d$  are linearly independent over  $\mathbb{Q}$ .

In order to provide some context for our results, let us focus for a moment on a specific transformation, namely, the doubling map  $x \mapsto 2x \bmod 1$ , which is a well-known ergodic transformation of the one-dimensional torus  $\mathbb{T}$ . The sum  $\sum_{n=0}^{N-1} f(2^n x)$  satisfies the central limit theorem and the law of iterated logarithm for a large class of functions. See the work of Fortet [For40], Kac [Kac46], and Maruyama [Mar50]. For subsequent extension of these results we mention, among others, the works of Aistleitner [Ais10, Ais13] and refer to the references therein. More in detail, Maruyama, building upon the results of Kac, proved that if  $f$  is a continuous function with vanishing mean and satisfying a Hölder condition of order  $\alpha > 0$ , then, for almost every  $x$ ,

$$\limsup_{N \rightarrow +\infty} \frac{1}{\sqrt{2N \log \log(N)}} \sum_{n=0}^{N-1} f(2^n x) = \lim_{N \rightarrow +\infty} \left( \frac{1}{N} \int_{\mathbb{T}} \left( \sum_{n=0}^{N-1} f(2^n y) \right)^2 dy \right)^{\frac{1}{2}}.$$

The point of view in the papers, we mentioned focuses on the lacunarity of the sequence  $\{2^n x\}_{n \in \mathbb{N}}$  and on the analogy with systems of independent random variables. In this work, instead, we take advantage of the fact that the composition operator  $Tf(x) = f(2x)$  is a shift operator on  $L^2(\mathbb{T}, dx)$  (see below for the exact definition).

Before stating our results, we briefly recall some definitions following [SNFBK10]. Let  $\mathcal{H}$  be a complex separable Hilbert space endowed with the inner product  $\langle \cdot, \cdot \rangle$ . Let  $T : \mathcal{H} \rightarrow \mathcal{H}$  be an isometry, that is, a bounded linear operator such that

$$\langle Tf, Tg \rangle = \langle f, g \rangle \quad \forall f, g \in \mathcal{H}.$$

A subspace  $\mathcal{V} \subseteq \mathcal{H}$  is called a *wandering subspace* for the isometry  $T : \mathcal{H} \rightarrow \mathcal{H}$  if

$$T^m(\mathcal{V}) \perp T^n(\mathcal{V}) \quad \forall m, n \in \mathbb{N} \cup \{0\}, m \neq n.$$

The isometry  $T : \mathcal{H} \rightarrow \mathcal{H}$  is a *unilateral shift* if there exists a wandering subspace  $\mathcal{V} \subseteq \mathcal{H}$  for  $T$  such that

$$\mathcal{H} = \bigoplus_{k \in \mathbb{N} \cup \{0\}} T^k(\mathcal{V}).$$

In this case, we say that the subspace  $\mathcal{V}$  is a *generating wandering subspace* for  $T$ . Notice that

$$\mathcal{V} = \mathcal{H} \ominus T(\mathcal{H}).$$

Unilateral shifts are ubiquitous in operator theory. One reason for this is provided by Wold’s decomposition theorem (see, e.g., [SNFBK10, Chapter 1]).

**Theorem 1.1** (Wold decomposition) *Let  $T : \mathcal{H} \rightarrow \mathcal{H}$  be an isometry. Then,*

$$\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^\perp,$$

where  $\mathcal{M}$  and  $\mathcal{M}^\perp$  are invariant under  $T$ ,  $T : \mathcal{M} \rightarrow \mathcal{M}$  is a unilateral shift and  $T : \mathcal{M}^\perp \rightarrow \mathcal{M}^\perp$  is a unitary operator. Such decomposition is uniquely determined and it holds

$$\mathcal{M} = \bigoplus_{k \in \mathbb{N} \cup \{0\}} T^k(\mathcal{H} \ominus T(\mathcal{H})), \quad \mathcal{M}^\perp = \bigcap_{k \in \mathbb{N}} T^k(\mathcal{H}).$$

Similarly to unilateral shifts, it is possible to define bilateral shifts. A subspace  $\mathcal{V} \subseteq \mathcal{H}$  is called a *wandering subspace* for the *unitary operator*  $T : \mathcal{H} \rightarrow \mathcal{H}$  if

$$T^m(\mathcal{V}) \perp T^n(\mathcal{V}) \quad \forall m, n \in \mathbb{Z}, m \neq n$$

and  $T : \mathcal{H} \rightarrow \mathcal{H}$  is a *bilateral shift* if there exists a generating wandering subspace  $\mathcal{V} \subseteq \mathcal{H}$  such that

$$\mathcal{H} = \bigoplus_{k \in \mathbb{Z}} T^k(\mathcal{V}).$$

Notice that for bilateral shifts the generating wandering subspace is not uniquely determined.

If  $T : \mathcal{H} \rightarrow \mathcal{H}$  is a shift, then  $\mathcal{H}$  admits an orthonormal basis of the form  $\{\varphi_{j,k}\}_{j \in \mathbb{X}, k \in \mathbb{Y}}$ , where  $\mathbb{X} \subseteq \mathbb{N}$  and  $\mathbb{Y}$  is either  $\mathbb{N} \cup \{0\}$  or  $\mathbb{Z}$  depending on  $T$  being a unilateral or bilateral shift, such that  $\{\varphi_{j,k}\}_{j \in \mathbb{X}}$  is an orthonormal basis for  $T^k(\mathcal{V})$  for every  $k \in \mathbb{Y}$  and such that, for every fixed  $k \in \mathbb{Y}$ , it holds

$$T\varphi_{j,k} = \varphi_{j,k+1}.$$

From now on when, we say that the isometry  $T : \mathcal{H} \rightarrow \mathcal{H}$  is a shift we mean that  $T$  could be either a unilateral or a bilateral shift. However, the reader has to keep in mind that whenever  $T$  is intended as a bilateral shift then  $T$  is not only an isometry, but a unitary operator as well.

We now introduce the general setting in which our results take place. We will assume the following:

- (i)  $\mathcal{H}$  is a Hilbert space and  $T : \mathcal{H} \rightarrow \mathcal{H}$  is an isometry.
- (ii)  $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^\perp$ , where  $T|_{\mathcal{M}} : \mathcal{M} \rightarrow \mathcal{M}$  is a shift (bilateral or unilateral) and  $T|_{\mathcal{M}^\perp} : \mathcal{M}^\perp \rightarrow \mathcal{M}^\perp$  is the identity operator; i.e., we are considering isometries whose unitary part in the Wold decomposition is the identity operator.
- (iii)  $\mathcal{V}$  is a generating wandering subspace for  $T|_{\mathcal{M}}$  and  $\Pi_{\mathcal{M}^\perp}$  and  $\Pi_k$  are the orthogonal projections from  $\mathcal{H}$  onto  $\mathcal{M}^\perp$  and  $T^k(\mathcal{V})$ , respectively. Here,  $k$  varies either in  $\mathbb{N} \cup \{0\}$  or  $\mathbb{Z}$  accordingly with the fact that  $T$  is a unilateral or a bilateral shift.

The following theorem is implicit in the existing literature, but we could not find a precise reference. In particular, when  $T$  is a shift such that  $\dim(\mathcal{V}) = +\infty$ , the theorem is proved in [Kre79] and [KP81]. Anyhow, a short proof will be included for the reader's convenience.

**Theorem 1.2** *With the notation above, for every positive vanishing sequence  $\varepsilon_n \rightarrow 0$  as  $n \rightarrow +\infty$ , there exists  $f \in \mathcal{H}$  such that*

$$\limsup_{N \rightarrow +\infty} \varepsilon_N^{-1} \left\| \frac{1}{N} \sum_{n=0}^{N-1} T^n f - \Pi_{\mathcal{M}^\perp} f \right\|_{\mathcal{H}} = +\infty.$$

Despite the negative result in the previous theorem, it is possible to give some positive results on the speed convergence under appropriate assumptions on the operator and on the functions. The following result is no surprising and we include it for the sake of completeness.

**Theorem 1.3** *With the notation above,*

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} T^n f - \Pi_{\mathcal{M}^\perp} f \right\|_{\mathcal{H}} \leq \frac{1}{\sqrt{N}} \sum_k \|\Pi_k f\|_{\mathcal{H}}.$$

Moreover, the rate of convergence  $1/\sqrt{N}$  is sharp.

The next theorem is our first main one. We obtain a result on the pointwise speed of convergence and the boundedness of a maximal function.

**Theorem 1.4** *With the notation above, assume that  $\mathcal{H}$  is the function space  $L^2_\mu := L^2(X, d\mu)$ , where  $(X, \mu)$  is a probability space, and that  $\varepsilon : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is a positive decreasing function. Define the maximal operator*

$$Sf(x) = \sup_{N \geq 1} N\varepsilon(N) \left| \frac{1}{N} \sum_{n=0}^{N-1} T^n f(x) - \Pi_{\mathcal{M}^\perp} f(x) \right|.$$

Then, there exists a positive constant  $c$  such that

$$(1.1) \quad \|Sf\|_{L^2(X, \mu)} \leq c \left( \sum_{n=0}^{+\infty} \varepsilon^2(n) \log^2(n+2) \right)^{\frac{1}{2}} \sum_k \|\Pi_k f\|_{L^2_\mu}.$$

Moreover, if

$$(1.2) \quad \sum_{n=0}^{+\infty} \varepsilon^2(n) \log^2(n+2) < +\infty \quad \text{and} \quad \sum_k \|\Pi_k f\|_{L^2_\mu} < +\infty,$$

then, for  $\mu$ -almost every  $x$ ,

$$(1.3) \quad \lim_{N \rightarrow +\infty} N\varepsilon(N) \left| \frac{1}{N} \sum_{n=0}^{N-1} T^n f(x) - \Pi_{\mathcal{M}^\perp} f(x) \right| = 0.$$

For example, one can choose  $\varepsilon(n) = n^{-\frac{1}{2}} \log^{-\delta}(n+2)$  with  $\delta > \frac{3}{2}$ . Then equation (1.3) gives a speed of convergence of the ergodic means of  $T$  at least of the order of  $N^{-\frac{1}{2}} \log^\delta(N+2)$ . Some particular instances of the above theorem, in the special

case that  $T$  is the operator of composition with a measure preserving transformation of  $X$ , have been obtained by Cuny [Cun11, Theorem 4.5] (see also Remark 1 after Theorem 4.5). The following are two straightforward applications of the above theorem. In Corollary 1.5, we consider functions defined on the square  $[0, 1]^2$  and their expansions with respect to the product Walsh system. We recall the definition of such system in the proof of the corollary. In Corollary 1.6, we consider the system of Laguerre polynomials, whose definition is, once again, recalled in the proof of the corollary. In both the corollaries, the almost every convergence is intended with respect to the Lebesgue measure.

**Corollary 1.5** Let  $B : [0, 1]^2 \rightarrow [0, 1]^2$  be the baker's transformation defined by

$$B(x, y) = \begin{cases} (2x, \frac{y}{2}), & \text{if } 0 \leq x < \frac{1}{2}, \\ (2x - 1, \frac{y}{2} + \frac{1}{2}), & \text{if } \frac{1}{2} \leq x < 1. \end{cases}$$

Assume that  $f$  has an absolutely convergent expansion with respect to the product Walsh system on the square  $[0, 1]^2$ . Then, for every  $\eta > 0$  and for almost every  $x$ ,

$$\lim_{N \rightarrow +\infty} \frac{\sqrt{N}}{(\log(1 + N))^{\frac{3}{2} + \eta}} \left| \frac{1}{N} \sum_{n=0}^{N-1} f(B^n x) - \int_{[0,1]^2} f(y) dy \right| = 0.$$

**Corollary 1.6** Let  $T$  be the operator

$$Tf(x) = f(x) - \int_0^x f(y) dy$$

defined on the Hilbert space  $L^2(\mathbb{R}_+, e^{-x} dx)$ , and let  $\{L_n\}_{n \in \mathbb{N}}$  be the system of Laguerre polynomials. Assume that the Laguerre coefficients of  $f$  are absolutely summable. Then, for every  $\eta > 0$  and for almost every  $x$ ,

$$\lim_{N \rightarrow +\infty} \frac{\sqrt{N}}{(\log(1 + N))^{\frac{3}{2} + \eta}} \left| \frac{1}{N} \sum_{n=0}^{N-1} T^n f(x) \right| = 0.$$

Our last theorem is about ergodic means associated with the endomorphisms of the two-dimensional torus  $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$  and the classical trigonometric expansion. We prove that it is enough to require a mild summability condition with respect to a logarithmic weight on the Fourier coefficients of a function to gain a speed of convergence essentially of order  $N^{-\frac{1}{2}}$  for the ergodic means.

**Theorem 1.7** Let  $A$  be a  $2 \times 2$  integer matrix such that  $\det(A) \neq 0$  and no eigenvalue of  $A$  is a root of unity. Assume that  $f \in L^2(\mathbb{T}^2, dx)$  has the trigonometric expansion

$$f(x) = \sum_{\xi \in \mathbb{Z}^2} \widehat{f}(\xi) e^{2\pi i x \xi}$$

and that, for some  $\delta > 0$ ,

$$(1.4) \quad \sum_{\xi \in \mathbb{Z}^2} (\log(1 + |\xi|))^{1+\delta} |\widehat{f}(\xi)|^2 < +\infty.$$

Then, for every  $\eta > 0$  and for almost every  $x \in \mathbb{T}^2$ ,

$$\lim_{N \rightarrow +\infty} \frac{\sqrt{N}}{(\log(1+N))^{\frac{3}{2}+\eta}} \left| \frac{1}{N} \sum_{n=0}^{N-1} f(A^n x) - \int_{\mathbb{T}^2} f(y) dy \right| = 0.$$

We point out that, in the above theorem,  $A$  has no eigenvalues which are not roots of unity if and only if  $A$  is an ergodic matrix [EW11, Corollary 2.2]. Therefore, the above theorem guarantees a speed of convergence for the ergodic means of a large class of functions for a particular instance of Birkhoff's ergodic theorem. Condition (1.4) is satisfied, for instance, by functions in any fractional Sobolev space. A more general sufficient condition in terms of the  $L^2$  integral modulus of continuity will be given in Proposition 4.1.

The situation in dimension  $d > 2$  seems to be more complicated. Nonetheless, we prove the following partial result, which is a corollary of Theorem 1.4.

**Corollary 1.8** *Let  $A$  be a  $d \times d$  matrix with integer coefficients and  $\det(A) \neq 0$ . Suppose there exists a set  $\mathcal{E} \subseteq \mathbb{Z}^d \setminus \{0\}$  such that the subspace of  $L^2_0(\mathbb{T}^d, dx)$*

$$\mathcal{V}_{\mathcal{E}} := \{f \in L^2_0(\mathbb{T}^d, dx) : \text{supp}(\widehat{f}) \subseteq \mathcal{E}\}$$

*is a generating wandering subspace for the operator  $T_A f = f \circ A$ . Suppose that there exist  $c > 0, q > 1$ , such that for all  $\xi \in \mathcal{E}$  and  $k \in \mathbb{Y}$  (where  $\mathbb{Y}$  is either  $\mathbb{N} \cup \{0\}$  or  $\mathbb{Z}$  depending on whether  $T_A$  is a unilateral or bilateral shift),*

$$(1.5) \quad |A^k \xi| \geq cq^{|k|}.$$

*Assume that  $f \in L^2(\mathbb{T}^2, dx)$  has the trigonometric expansion*

$$f(x) = \sum_{\xi \in \mathbb{Z}^2} \widehat{f}(\xi) e^{2\pi i x \xi}$$

*and that, for some  $\delta > 0$ ,*

$$(1.6) \quad \sum_{\xi \in \mathbb{Z}^2} (\log(1+|\xi|))^{1+\delta} |\widehat{f}(\xi)|^2 < +\infty.$$

*Then, for every  $\eta > 0$  and for almost every  $x \in \mathbb{T}^2$ ,*

$$\lim_{N \rightarrow +\infty} \frac{\sqrt{N}}{(\log(1+N))^{\frac{3}{2}+\eta}} \left| \frac{1}{N} \sum_{n=0}^{N-1} f(A^n x) - \int_{\mathbb{T}^2} f(y) dy \right| = 0.$$

Assumption (1.5) is satisfied, for instance, whenever  $A$  is an expansive matrix, i.e., whenever there exists  $q > 1$  such that  $|Ax| \geq q|x|$  for all  $x \in \mathbb{R}^d$ .

We should also mention that in the literature there exist theorems of flavor similar to Theorem 1.7. For example in [LÖ14, Theorem 1.2], the author proves the law of the iterated logarithm for averages of the form

$$\frac{1}{N} \sum_{n=0}^{N-1} f(M_n x),$$

where  $(M_n)_{n \geq 1}$  is a sequence of integer matrices satisfies a strong Hadamard-type condition [LÖ14, Condition (1.4)] and  $f$  is a function of finite Hardy–Krause total

variation. Although our theorem gives less precise asymptotic information than the law of the iterated logarithm, our assumptions are much less stringent. If  $A$  is a matrix as in Theorem 1.7, then the sequence  $M_n := A^n$  does not in general satisfy [Lö14, Condition (1.4)] and functions satisfying (1.4) can be quite rough. Furthermore, for matrices with eigenvalues of modulus greater than 1, Fan [Fan99] has obtained sharp estimates for the decay of correlation which lead to central limit-type theorems for the distribution of values of the ergodic averages.

## 2 Proof of Theorems 1.2, 1.3, and 1.4 and of Corollaries 1.5 and 1.6

The proof of Theorem 1.2 is straightforward.

**Proof of Theorem 1.2** Since  $T$  has operator norm 1, the averaging operator  $U_N := \frac{1}{N} \sum_{n=0}^{N-1} T^n$  has operator norm at most 1. Furthermore, the norm is at least 1, as it can be seen by testing the operator  $U_N$  on the functions  $f_H = \sum_{k=0}^H \varphi_{j,k}$  and letting  $H \rightarrow +\infty$ . Here,  $\{\varphi_{j,k}\}_{j,k}$  is an orthonormal basis associated with the shift  $T$ . Therefore, the family of operators  $\{\varepsilon_N^{-1} U_N\}_N$  is not uniformly bounded in the operator norm. Hence, by the Banach–Steinhaus uniform boundedness principle, there exists  $f \in \mathcal{M} \subseteq \mathcal{H}$  such that

$$\limsup_{N \rightarrow +\infty} \varepsilon_N^{-1} \left\| \frac{1}{N} \sum_{n=0}^{N-1} T^n f \right\|_{\mathcal{H}} = +\infty. \quad \blacksquare$$

As mentioned, Theorem 1.3 can also be proved using the unitary equivalence with the shift operator on vector valued Hardy spaces in the unit disc. However, for the sake of completeness, we provide here a direct proof.

**Proof of Theorem 1.3** The proof for unilateral or bilateral shifts is the same. Let  $T : \mathcal{M} \rightarrow \mathcal{M}$  be a bilateral shift. Then, there exists a generating wandering subspace  $\mathcal{V}$  such that

$$\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^\perp = \left( \bigoplus_{k \in \mathbb{Z}} T^k(\mathcal{V}) \right) \oplus \mathcal{M}^\perp.$$

Let  $\{\varphi_{j,k}\}_{j \in \mathbb{X}, k \in \mathbb{Z}}$  be an orthonormal basis of  $\mathcal{M}$  associated with  $T$ . Without losing generality, we assume that  $f$  has only finitely many nonzero coefficients  $\{\widehat{f}(j,k)\}_{j \in \mathbb{X}, k \in \mathbb{Z}}$  with respect to the orthonormal basis  $\{\varphi_{j,k}\}_{j \in \mathbb{X}, k \in \mathbb{Z}}$ . Since  $T$  acts as the identity on  $\mathcal{M}^\perp$ , we have

$$\frac{1}{N} \sum_{n=0}^{N-1} T^n f - \Pi_{\mathcal{M}^\perp} f = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{j,k} \widehat{f}(j,k) \varphi_{j,k+n} = \frac{1}{\sqrt{N}} \sum_{j,k} \widehat{f}(j,k) \Psi_{j,k}(N),$$

where we have set

$$\Psi_{j,k}(N) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \varphi_{j,k+n}.$$

It can be readily checked that  $\{\Psi_{j,k}(N)\}_{j \in \mathbb{X}}$  is an orthonormal system for every fixed  $k \in \mathbb{Z}$ . Hence, by Parseval's identity,

$$\begin{aligned} \left\| \frac{1}{N} \sum_{n=0}^{N-1} T^n f - \Pi_{\mathcal{M}^\perp} f \right\| &= \left\| \frac{1}{\sqrt{N}} \sum_{j,k} \widehat{f}(j,k) \Psi_{j,k}(N) \right\| \\ &\leq \frac{1}{\sqrt{N}} \sum_{k \in \mathbb{Z}} \left\| \sum_{j \in \mathbb{X}} \widehat{f}(j,k) \Psi_{j,k}(N) \right\| \\ &= \frac{1}{\sqrt{N}} \sum_{k \in \mathbb{Z}} \left( \sum_{j \in \mathbb{X}} |\widehat{f}(j,k)|^2 \right)^{\frac{1}{2}} = \frac{1}{\sqrt{N}} \sum_{k \in \mathbb{Z}} \|\Pi_k f\|. \end{aligned}$$

Finally, observe that if  $f = \Pi_k f$  for a single  $k$ , then all the above inequalities actually are identities. Hence, the theorem is sharp. ■

The proof of Theorem 1.4 is in principle similar to the proof of Theorem 1.3. The main ingredient is the Rademacher–Menshov theorem, which we now recall.

**Theorem 2.1** (Rademacher–Menshov) *There exists an absolute positive constant  $C$  such that for every positive measure space  $(X, \mu)$  and every orthogonal system  $f_0, f_1 \dots$  in  $L^2(X, \mu)$ , the maximal function*

$$\mathcal{M}(x) := \sup_{k \geq 0} \left| \sum_{n=0}^k f_n(x) \right|$$

satisfies the estimate

$$\|\mathcal{M}\|_{L^2(X, \mu)} \leq C \left( \sum_{n=0}^{+\infty} \log^2(n+2) \|f_n\|_{L^2(X, \mu)}^2 \right)^{\frac{1}{2}}.$$

It is important to emphasize that the constant  $C$  in the above theorem is absolute and we refer the reader to [Mea07] for a discussion on this.

Recall also the next lemma by Kronecker, which is an application of Abel's summation by parts formula.

**Lemma 2.2** *Suppose that  $a_n$  is a sequence of complex numbers such that  $\sum_{n=1}^\infty a_n$ , exists and is finite. Assume also that  $b_n$  is a nondecreasing sequence of positive numbers tending to infinity. Then,*

$$\lim_{N \rightarrow \infty} \frac{1}{b_N} \sum_{n=0}^{N-1} b_n a_n = 0.$$

**Proof of Theorem 1.4** We assume again that  $T$  is a bilateral shift. The proof for the unilateral case is the same. To simplify the notation, we also assume that  $f$  is in  $\mathcal{M}$ , so that  $\Pi_{\mathcal{M}^\perp} f = 0$ . Finally, assume that  $f$  has only finitely many nonzero Fourier coefficients. Then,

$$N\varepsilon(N) \left( \frac{1}{N} \sum_{n=0}^{N-1} T^n f(x) - \Pi_{\mathcal{M}^\perp} f(x) \right) = \varepsilon(N) \sum_{n=0}^{N-1} T^n f(x).$$



We derive both (1.1) and (1.3) from the boundedness of an auxiliary maximal function. Let  $\varepsilon : [0, +\infty) \rightarrow \mathbb{R}$ , not necessarily decreasing, and define

$$\tilde{S}f(x) := \sup_{N \geq 1} \left| \sum_{n=0}^{N-1} \varepsilon(n) T^n f(x) \right|.$$

We have

$$\sum_{n=0}^{N-1} \varepsilon(n) T^n f(x) = \sum_{n=0}^{N-1} \varepsilon(n) \sum_{j \in \mathbb{X}, k \in \mathbb{Z}} \widehat{f}(j, k) \varphi_{j, k+n}(x) = \sum_{k \in \mathbb{Z}} A(k) \sum_{n=0}^{N-1} \varepsilon(n) \Phi(k, n, x),$$

where we have set

$$A(k) = \|\Pi_k f\|_{L^2_\mu} = \left( \sum_{j \in \mathbb{X}} |\widehat{f}(j, k)|^2 \right)^{\frac{1}{2}}, \quad \Phi(k, n, x) = \frac{1}{A(k)} \sum_{j \in \mathbb{X}} \widehat{f}(j, k) \varphi_{j, k+n}(x).$$

Then,

$$\sup_{N \geq 1} \left| \sum_{n=0}^{N-1} \varepsilon(n) T^n f(x) \right| = \sup_{N \geq 1} \left| \sum_{k \in \mathbb{Z}} A(k) \sum_{n=0}^{N-1} \varepsilon(n) \Phi(k, n, x) \right|.$$

In the above formula, we simply omit the terms such that  $A(k) = 0$ . It may be promptly verified that  $\{\Phi(k, n, x)\}_{n=0}^{N-1}$  is an orthonormal system for every fixed  $k \in \mathbb{Z}$  and  $N \in \mathbb{N}$ . Hence, by means of the Rademacher–Menshov theorem,

$$\begin{aligned} \|\tilde{S}f\|_{L^2_\mu} &= \left( \int_X \left( \sup_{N \geq 1} \left| \sum_{n=0}^{N-1} \varepsilon(n) T^n f(x) \right| \right)^2 d\mu(x) \right)^{\frac{1}{2}} \\ &\leq \sum_{k \in \mathbb{Z}} A(k) \left( \int_X \left( \sup_{N \geq 1} \left| \sum_{n=0}^{N-1} \varepsilon(n) \Phi(k, n, x) \right| \right)^2 d\mu(x) \right)^{\frac{1}{2}} \\ (2.1) \quad &\leq c \left( \sum_{n \in \mathbb{N}} \varepsilon^2(n) \log^2(n+2) \right)^{\frac{1}{2}} \sum_{k \in \mathbb{Z}} \|\Pi_k f\|_{L^2_\mu}. \end{aligned}$$

Now a standard argument, as in [Zyg03, p. 190], shows that inequality (2.1) with condition (1.2) implies that the series  $\sum_{n=0}^\infty \varepsilon(n) T^n f(x)$  converges  $\mu$ -a.e. Moreover, restricting to a positive decreasing  $\varepsilon$ , we apply Kronecker’s lemma with  $a_n = \varepsilon(n) T^n f(x)$ ,  $b_n = \varepsilon^{-1}(n)$  and we have that

$$\lim_{N \rightarrow \infty} \varepsilon(N) \left| \sum_{n=0}^{N-1} T^n f(x) \right| = 0, \quad \mu\text{-a.e.},$$

which proves (1.3).

In order to prove (1.1), assume again that  $\varepsilon$  is positive and decreasing. Then, by Abel’s summation by parts,

$$\begin{aligned} \varepsilon(N) \sum_{n=0}^{N-1} T^n f(x) &= \frac{\varepsilon(N)}{\varepsilon(N-1)} \sum_{n=0}^{N-1} \varepsilon(n) T^n f(x) \\ &\quad - \varepsilon(N) \sum_{j=0}^{N-2} \left( \sum_{n=0}^j \varepsilon(n) T^n f(x) \right) \left( \frac{1}{\varepsilon(j+1)} - \frac{1}{\varepsilon(j)} \right). \end{aligned}$$

Hence,

$$(2.2) \quad Sf(x) = \sup_{N \geq 1} \varepsilon(N) \left| \sum_{n=0}^{N-1} T^n f(x) \right| \leq 2 \sup_{N \geq 1} \left| \sum_{n=0}^{N-1} \varepsilon(n) T^n f(x) \right| = 2 \tilde{S}f(x).$$

This, together with (2.1) proves (1.1). ■

We conclude the section showing that the hypothesis of Theorem 1.4 are satisfied in the setting of Corollaries 1.5 and 1.6.

**Proof of Corollary 1.5** One can verify that the composition operator  $T_B f(x, y) = f(B(x, y))$  is a bilateral shift with respect to the product Walsh system on the square  $[0, 1]^2$ , whose definition we now recall. Let  $r_k$  be the one-dimensional  $k$ th Rademacher function

$$r_k(x) = \text{sgn}(\sin(2^k \pi x)), \quad k \in \mathbb{N}, x \in [0, 1).$$

On the unit square  $[0, 1]^2$  define the function

$$R_k(x, y) := \begin{cases} r_{k+1}(x) & k = 0, 1, 2, \dots \\ r_{|k|}(y) & k = -1, -2, \dots \end{cases}$$

and for every set of integers  $k_1 < k_2 < \dots < k_n$  define

$$W_{k_1 k_2 \dots k_n}(x, y) = R_{k_1}(x, y) \dots R_{k_n}(x, y).$$

Then,

$$L_0^2([0, 1]^2) = \overline{\text{span}} \left\{ W_{k_1 k_2 \dots k_n} : k_1 < k_2 < \dots < k_n, k_j \in \mathbb{Z}, n \in \mathbb{N} \right\},$$

where  $L_0^2([0, 1]^2)$  is the subspace of  $L^2([0, 1]^2)$  consisting of functions with vanishing mean. One can verify that

$$T(W_{k_1 k_2 \dots k_n}) = W_{(k_1+1)(k_2+1) \dots (k_n+1)}.$$

Hence, the transformation  $T$  is a bilateral shift on  $L_0^2([0, 1]^2)$  with a generating wandering subspace given by

$$\mathcal{V} = \overline{\text{span}} \left\{ W_{1k_2 \dots k_n}, 1 < k_2 < k_3 < \dots < k_n, k_j \in \mathbb{Z}, n \in \mathbb{N} \right\}.$$

Then, Theorem 1.4 applies. ■

**Proof of Corollary 1.6** Recall the definition of Laguerre polynomials  $\{L_n\}_{n \in \mathbb{N}}$ ,

$$L_n(x) = \frac{e^x}{n!} \frac{d^n}{dx^n} (e^{-x} x^n) = \sum_{k=0}^n \binom{n}{k} \frac{(-1)^k}{k!} x^k.$$

This family of polynomials is an orthonormal basis for the Hilbert space  $L^2(\mathbb{R}_+, e^{-x} dx)$ . As observed by Von Neumann [vN29] (see also Brown and Halmos [BHS65, p. 135]), the operator

$$Tf(x) = f(x) - \int_0^x f(y) dy$$

is the unilateral shift with respect to the Laguerre basis of  $L^2(\mathbb{R}_+, e^{-x} dx)$ . Indeed,

$$TL_n(x) = \sum_{k=0}^n \binom{n}{k} \frac{(-1)^k}{k!} x^k + \sum_{k=0}^n \binom{n}{k} \frac{(-1)^{k+1}}{(k+1)!} x^{k+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} \frac{(-1)^k}{k!} x^k = L_{n+1}(x).$$

Hence, Theorem 1.4 applies. ■

### 3 Speed of convergence for toral endomorphisms

Before actually proving Theorem 1.7 and Corollary 1.8, we make some preliminary observations. If in Theorem 1.7, we choose a matrix  $A$  with  $|\det A| > 1$ , then the operator  $T_A f = f \circ A$  is a unilateral shift on  $L^2_0(\mathbb{T}^2)$ , the space of square integrable functions with vanishing means. This is proved, e.g., in [Krz93], but it will also follow from the proof of Lemma 3.3. If, on the other hand,  $|\det A| = 1$ , then  $T_A$  is a bilateral shift on  $L^2_0(\mathbb{T}^2)$ . A generating wandering subspace for  $T_A$  can be constructed as follows. Let us consider equivalence relation on  $\mathbb{Z}^2 \setminus \{0\}$  defined by the orbits of  $A^*$ , i.e.,

$$\xi \sim \mu \iff \exists k \in \mathbb{Z} : A^{*k} \xi = \mu.$$

Let now  $\mathcal{E}$  be the set containing of representative from each equivalence class of  $\mathbb{Z}^2 \setminus \{0\} / \sim$ . A generating wandering subspace  $\mathcal{V}_{\mathcal{E}}$  for  $T_A$  is then given by

$$(3.1) \quad \mathcal{V}_{\mathcal{E}} = \{f \in L^2(\mathbb{T}^d) : \text{supp } \widehat{f} \subseteq \mathcal{E}\}.$$

The proof of Theorem 1.7 will follow from a series of preparatory results. In particular, we deal with the cases  $|\det A| > 1$  and  $|\det A| = 1$  in different ways. In this latter case, we will have to be more careful in constructing a generating wandering subspace  $\mathcal{V}_{\mathcal{E}}$ , which we recall is not unique for bilateral shifts.

#### 3.1 Proof of Theorem 1.7: case $|\det A| = 1$

Let  $\text{tr}(A)$  be the trace of the matrix  $A$ . Observe that if  $\det(A) = 1$  the eigenvalues of  $A$  are given by

$$\frac{\text{tr}(A) \pm \sqrt{\text{tr}^2(A) - 4}}{2}.$$

Since no eigenvalue of  $A$  is a root of unity by hypothesis, we can assume that  $|\text{tr}(A)| > 2$ . Otherwise, that is, if  $\text{tr}(A) = 0, \pm 1, \pm 2$ , it can be checked by hand that the eigenvalues of  $A$  are roots of unity and in this case Birkhoff's theorem would not apply since the matrix  $A$  would not be ergodic (see [Krz93]). If  $\det(A) = -1$ , then the eigenvalues of  $A$  are given by

$$\frac{\text{tr}(A) \pm \sqrt{\text{tr}^2(A) + 4}}{2}.$$

Notice that these are roots of unity if and only if  $\text{tr}(A) = 0$ . In all remaining cases, we have two distinct eigenvalues  $\lambda, \lambda^{-1} \in \mathbb{R}$  and, without loss of generality, we can assume

that  $0 < |\lambda|^{-1} < 1 < |\lambda|$ . We now take advantage of this to define a suitable generating wandering subspace for the bilateral shift  $T_A$ . Let  $S \in GL_2(\mathbb{R})$  be such that

$$A = S^{-1} \begin{bmatrix} \lambda^{-1} & 0 \\ 0 & \lambda \end{bmatrix} S = S^{-1} D S.$$

Let  $\mathcal{E} \subseteq \mathbb{Z}^2 \setminus \{0\}$  be such that it contains exactly one element from each orbit of the action of  $A$  on  $\mathbb{Z}^2$ . We choose such element as follows. Define  $|\xi|_\infty = |(\xi_1, \xi_2)|_\infty = \max\{|\xi_1|, |\xi_2|\}$ . Let  $\mathcal{O}$  an orbit of  $A$  in  $\mathbb{Z}^2 \setminus \{0\}$  and consider the set  $S\mathcal{O}$ . Then, we choose  $\xi \in \mathcal{O}$  such that  $S\xi$  has the minimal  $|\cdot|_\infty$  norm. Equivalently, for all  $k \in \mathbb{Z}$ , we have that

$$(3.2) \quad |SA^k \xi|_\infty = |D^k S \xi|_\infty \geq |S \xi|_\infty.$$

Then, a generating wandering subspace for  $T_A$  is defined as in (3.1).

Using the notation above, we prove the following.

**Lemma 3.1** *Let  $A$  be a  $2 \times 2$  integer matrix such that  $|\det A| = 1$  and no eigenvalues of  $A$  is a root of unity. Let  $\mathcal{E}$  be defined as above. Then, there exist constants  $c > 0$  and  $q > 1$  such that, for every  $k \in \mathbb{Z}$ ,*

$$\min\{|A^k \xi| : \xi \in \mathcal{E}\} \geq cq^{|k|}.$$

**Proof** Assume that  $\det A = 1$ ; the case  $\det A = -1$  is similar. Since for every  $\xi \in \mathbb{Z}^2 \setminus \{0\}$  and  $k \in \mathbb{Z}$  it holds that

$$|A^k \xi| = |S^{-1} D^k S \xi| \geq \|S\|^{-1} |D^k S \xi|,$$

and all norms in a finite-dimensional vector space are equivalent, it suffices to show that there exist  $c > 0, q > 1$  such that  $|D^k S \xi|_\infty \geq cq^{|k|}$  for every  $\xi \in \mathcal{E}$ . Let  $\eta = (\eta_1, \eta_2) = S\xi$  where  $\xi$  is in  $\mathcal{E}$  and let  $\lambda^{-1}, \lambda$  the two real eigenvalues of  $A$  with  $|\lambda| > 1$ . Then,

$$|D^k \eta|_\infty = |(\lambda^{-1} \eta_1, \lambda^k \eta_2)|_\infty \geq |\lambda|^k |\eta_2| \geq \min\{|\eta_1|, |\eta_2|\} |\lambda|^k$$

and, similarly,

$$|D^k \eta|_\infty = |(\lambda^{-1} \eta_1, \lambda^k \eta_2)|_\infty \geq |\lambda|^{-k} |\eta_1| \geq \min\{|\eta_1|, |\eta_2|\} |\lambda|^{-k}.$$

Hence,

$$(3.3) \quad |D^k \eta|_\infty \geq \min\{|\eta_1|, |\eta_2|\} |\lambda|^{|k|}.$$

The conclusion will follow once we prove that  $\min\{|\eta_1|, |\eta_2|\}$  is bounded from below uniformly for  $\eta = (\eta_1, \eta_2)$  in  $\mathcal{E}$ . But this is true because of the following. If  $|\eta_2| \leq |\eta_1|$ , by the definition of  $\mathcal{E}$ ,

$$|(\eta_1, \eta_2)|_\infty \leq |D(\eta_1, \eta_2)|_\infty = |(\lambda^{-1} \eta_1, \lambda \eta_2)|_\infty = |\lambda| |\eta_2|.$$

The last identity holds since if  $|(\lambda^{-1} \eta_1, \lambda \eta_2)|_\infty = |\lambda^{-1}| |\eta_1|$ , then we would have  $|\eta_1| \leq |\lambda^{-1}| |\eta_1|$ , which is a contradiction since  $|\lambda^{-1}| < 1$  and  $\eta_1 \neq 0$ . Similarly, if  $|\eta_1| \leq |\eta_2|$ ,

$$|(\eta_1, \eta_2)|_\infty \leq |D^{-1}(\eta_1, \eta_2)|_\infty = |(\lambda \eta_1, \lambda^{-1} \eta_2)|_\infty = |\lambda| |\eta_1|.$$

Hence,  $|(\eta_1, \eta_2)|_\infty \leq |\lambda| \min\{|\eta_1|, |\eta_2|\}$ , that is,  $|\eta_1|$  and  $|\eta_2|$  are comparable. Therefore, by (3.3),

$$|D^k \eta|_\infty \geq \min\{|\eta_1|, |\eta_2|\} |\lambda|^{|k|} \geq |\eta|_\infty |\lambda|^{k-1} \geq c |\lambda|^k$$

for some positive constant  $c$ . This follows from the fact that  $\eta \in \mathbb{S}\mathbb{Z}^2 \setminus \{0\}$ . ■

We now conclude the proof of Theorem 1.7 in the case  $|\det A| = 1$ . As observed at the beginning of Section 3, the operator  $T_A f(x) = f(Ax)$  is a bilateral shift on  $L^2_0(\mathbb{T}^2)$  with a generating subspace given by  $\mathcal{V}_\mathcal{E}$  as in (3.1) where  $\mathcal{E}$  is defined by means of the property (3.2). Hence, Theorem 1.4 applies and, in particular, it applies with  $\varepsilon(n) = (n + 1)^{-\frac{1}{2}} (\log(2 + n))^{-\frac{3}{2} - \eta}$  for any  $\eta > 0$ .

Set now  $\mathcal{F}_k := (A^*)^k \mathcal{E}$ . Observe that  $A$  satisfies the hypothesis of Lemma 3.1 if and only if  $A^*$  does. Hence, by such lemma, there exist constants  $c > 0$  and  $q > 1$  such that, for every  $k \in \mathbb{Z}$ ,

$$\min\{|\xi| : \xi \in (A^*)^k \mathcal{E} \setminus \{0\}\} \geq cq^{|k|}.$$

Hence, for every positive increasing function  $v$  and  $f$  satisfying (1.4), one has

$$\begin{aligned} \sum_{k \in \mathbb{Z} \cup \{0\}} \|\Pi_k f\|_{L^2} &= \sum_{k \in \mathbb{Z} \cup \{0\}} \left( \sum_{\xi \in \mathcal{F}_k} |\widehat{f}(\xi)|^2 \right)^{\frac{1}{2}} \\ &\leq \left( \sum_{k \in \mathbb{Z} \cup \{0\}} v^{-2}(k) \right)^{\frac{1}{2}} \left( \sum_{k \in \mathbb{Z} \cup \{0\}} v^2(k) \sum_{\xi \in \mathcal{F}_k} |\widehat{f}(\xi)|^2 \right)^{\frac{1}{2}} \\ &\leq \left( \sum_{k \in \mathbb{Z} \cup \{0\}} v^{-2}(k) \right)^{\frac{1}{2}} \left( \sum_{k \in \mathbb{Z} \cup \{0\}} \sum_{\xi \in \mathcal{F}_k} v^2 \left( \frac{\log |\xi| - \log c}{\log q} \right) |\widehat{f}(\xi)|^2 \right)^{\frac{1}{2}}. \end{aligned}$$

The conclusion follows choosing  $v(t) = t^{\frac{1}{2} + \frac{\delta}{2}} + 1$ .

### 3.2 Proof of Theorem 1.7: case $|\det A| > 1$

We want to prove the analogous of Lemma 3.1 for a matrix  $A$  with  $|\det A| > 1$ . However, we need a preliminary result, which is a special case of [Kat71, Lemma 3]. The proof we provide here for the reader's convenience is essentially the same one as in [Kat71] adapted to the case  $d = 2$ .

**Lemma 3.2** *Let  $A$  be a  $2 \times 2$  integer matrix with a real irrational eigenvalue  $\lambda$ , and let  $V_\lambda$  be its corresponding eigenspace. Then, there exists  $C_A > 0$  such that, for  $\xi \in \mathbb{Z}^2 \setminus \{0\}$ ,*

$$|\xi| \operatorname{dist}(\xi, V_\lambda) \geq C_A,$$

where  $\operatorname{dist}$  is the Euclidean distance between  $\xi$  and  $V_\lambda$ .

**Proof** By Dirichlet's theorem, for every  $Q \in \mathbb{N}$ , there exists  $q \in \mathbb{N}$ ,  $q \leq Q$  and  $r \in \mathbb{Z}$  such that

$$\left| \lambda - \frac{r}{q} \right| < \frac{1}{qQ}.$$

Now, fix  $\xi \in \mathbb{Z}^2 \setminus \{0\}$  and notice that  $(qA - r)\xi \in \mathbb{Z}^2 \setminus \{0\}$ , so  $1/q \leq |(A - r/q)\xi|$ . Let  $y$  be the orthogonal projection of  $\xi$  on  $V_\lambda$ . We have

$$\begin{aligned} \frac{1}{q} &\leq \left| \left( A - \frac{r}{q} \right) \xi \right| = \left| \left( A - \frac{r}{q} \right) (\xi - y) + \left( \lambda - \frac{r}{q} \right) y \right| \\ &\leq (\|A\| + |\lambda| + 1) \operatorname{dist}(\xi, V_\lambda) + \frac{|\xi|}{qQ}. \end{aligned}$$

Setting  $C = \|A\| + |\lambda| + 1$  and rearranging the above inequality, we get

$$\left( 1 - \frac{|\xi|}{Q} \right) \leq C \operatorname{dist}(\xi, V_\lambda) q \leq C \operatorname{dist}(\xi, V_\lambda) Q.$$

Setting  $Q = \lceil 2|\xi| \rceil$  we obtain the desired estimate. ■

**Lemma 3.3** *Let  $A$  be a  $2 \times 2$  integer matrix such that  $|\det A| > 1$  and no eigenvalue of  $A$  is a root of unity. Then, there exist constants  $c > 0$  and  $q > 1$  such that, for every  $k \in \mathbb{N}$*

$$\min\{|\xi| : \xi \in A^k \mathbb{Z}^2 \setminus \{0\}\} \geq cq^k.$$

**Proof** We study separately the cases when  $A$  is diagonalizable and when it is not. Denote by  $\lambda, \Lambda \in \mathbb{C}$  the eigenvalues of the matrix  $A$  so that  $|\lambda| \leq |\Lambda|$ . Recall that  $\det(A) = \lambda\Lambda$  is an integer different from  $-1, 1, 0$ . If these eigenvalues are complex, then they are conjugate to each other and  $1 < |\lambda| = |\Lambda|$ . If the eigenvalues are real, then, either  $1 < |\lambda| \leq |\Lambda|$  or  $|\lambda| < 1 < |\Lambda|$ . In this last case,  $\lambda$  and  $\Lambda$  cannot be rational, since the characteristic polynomial of  $A$  is a monic polynomial with integer coefficients and any rational root of such polynomial is an integer.

$A$  is diagonalizable and  $1 < |\lambda| \leq |\Lambda|$ . In this case, there exists  $S \in \operatorname{GL}_2(\mathbb{C})$  such that for every  $k \in \mathbb{N}$

$$A^k = S^{-1} \begin{bmatrix} \lambda^k & 0 \\ 0 & \Lambda^k \end{bmatrix} S.$$

Therefore, for  $\xi \in \mathbb{Z}^2 \setminus \{0\}$ ,

$$|A^k \xi| \geq \frac{1}{\|S\|} \left\| \begin{bmatrix} \lambda^k & 0 \\ 0 & \Lambda^k \end{bmatrix} S\xi \right\| \geq \frac{|\lambda|^k |\xi|}{\|S\| \|S^{-1}\|} \geq \frac{|\lambda|^k}{\|S\| \|S^{-1}\|},$$

and the claim is proved in this case.

$A$  is diagonalizable and  $|\lambda| < 1 < |\Lambda|$ . Let  $V_\lambda, V_\Lambda$  be the one-dimensional eigenspaces corresponding to  $\lambda$  and  $\Lambda$ , respectively, let  $\theta \in (0, \pi)$  be the angle between them, and let  $P_\lambda, P_\Lambda$  be the oblique projections with respect to the axes  $V_\lambda, V_\Lambda$ . Define a new norm in  $\mathbb{R}^2$  as follows,

$$[\xi]_A := |P_\lambda \xi| + |P_\Lambda \xi|.$$

This is of course equivalent to the Euclidean norm of  $\mathbb{R}^2$  up to multiplicative constants which depends on  $A$ . In what follows,  $c, C$  denote positive constants which depend only on  $A$  and might change from appearance to appearance. Applying now Lemma 3.2 for some  $\xi \in \mathbb{Z}^2 \setminus \{0\}$ , we have

$$\operatorname{dist}(\xi, V_\lambda) = |\sin(\theta)| |P_\Lambda \xi| \geq c |\xi|^{-1} \geq c [\xi]_A^{-1}.$$

Hence,

$$|P_\Lambda \xi| \geq c[\xi]_A^{-1}.$$

Writing  $\xi = P_\lambda \xi + P_\Lambda \xi$  and applying  $A^k$ , we obtain  $A^k \xi = \lambda^k P_\lambda \xi + \Lambda^k P_\Lambda \xi$ . Hence,

$$\begin{aligned} C|A^k \xi| &\geq [A^k x]_A = |\lambda|^k |P_\lambda \xi| + |\Lambda|^k |P_\Lambda \xi| \\ &= |\lambda|^k (|\xi| - |P_\Lambda \xi|) + |\Lambda|^k |P_\Lambda \xi| \\ &= |\lambda|^k [\xi]_A + (|\Lambda|^k - |\lambda|^k) |P_\Lambda \xi| \\ &\geq |\lambda|^k [\xi]_A + c \frac{|\Lambda|^k - |\lambda|^k}{[\xi]_A} =: f([\xi]_A), \end{aligned}$$

where  $f(t) = |\lambda|^k t + c(|\Lambda|^k - |\lambda|^k)t^{-1}$ ,  $t > 0$ . Such function  $f$  admits a global minimum at  $t_{\min}$ ,

$$t_{\min} = \sqrt{\frac{c(|\Lambda|^k - |\lambda|^k)}{|\lambda|^k}}, \quad f(t_{\min}) = 2\sqrt{c|\lambda|^k(|\Lambda|^k - |\lambda|^k)}.$$

For  $k$  sufficiently large the estimate  $f(t_{\min}) \geq c\sqrt{|\lambda|^k |\Lambda|^k} = c|\det(A)|^{\frac{k}{2}}$  holds true, and this, combined with the above estimate, proves the claim.

*A is not diagonalizable.* In this case, we have a single eigenvalue  $\lambda$  with  $2\lambda = \text{tr}(A)$  and  $\lambda^2 = \det(A)$ . Hence,  $\text{tr}(A)^2 = 4 \det(A)$ . This implies that  $\text{tr}(A)$  is an even integer, so that  $\lambda \in \mathbb{Z} \setminus \{0\}$ . The Jordan decomposition of  $A$  guarantees that

$$A = S^{-1} \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} S,$$

for some  $S \in GL_2(\mathbb{C})$ . However, since the columns of  $S$  are obtained by solving a homogeneous system of linear equations with integer coefficients, we can assume, without loss of generality, that  $S$  has integer entries. Assume for the moment that  $\lambda|k$ , i.e., there exists  $q \in \mathbb{Z}$  such that  $k = q\lambda$ . Then,

$$A^k = S^{-1} \begin{bmatrix} \lambda^k & k\lambda^{k-1} \\ 0 & \lambda^k \end{bmatrix} S = \lambda^k S^{-1} \begin{bmatrix} 1 & q \\ 0 & 1 \end{bmatrix} S.$$

Notice that  $U := \begin{bmatrix} 1 & q \\ 0 & 1 \end{bmatrix}$  is in  $GL_2(\mathbb{Z})$ . Therefore,

$$A^k \mathbb{Z}^2 = \lambda^k S^{-1} U S \mathbb{Z}^2 \subseteq \lambda^k S^{-1} U \mathbb{Z}^2 = \lambda^k S^{-1} \mathbb{Z}^2.$$

In particular, it follows that

$$\delta_k := \min \{ |y| : y \in A^k \mathbb{Z}^2 \setminus \{0\} \} \geq \|S\|^{-1} |\lambda|^k,$$

which proves the claim when  $\lambda|k$ . In general, let  $k \equiv r \pmod{|\lambda|}$ ,  $0 \leq r < |\lambda|$ . Then,  $\delta_n \geq \delta_{k-r} \geq c|\lambda|^{k-r} \geq (c|\lambda|^{-|\lambda|})|\det(A)|^{\frac{k}{2}}$ , and this concludes the proof for a nondiagonalizable matrix  $A$ . ■

We now conclude the proof of Theorem 1.7 in the case  $|\det A| > 1$ . Notice that  $A$  satisfies the hypothesis of Lemma 3.3 if and only if  $A^*$  does. By Wold's theorem, the unitary part of the operator  $T_A f = f \circ A$  acts on the subspace  $\bigcap_{k \in \mathbb{N} \cup \{0\}} T_A^k(L_0^2(\mathbb{T}^2))$ , but this intersection is trivial and this follows at once from the fact that  $\bigcap_{k \in \mathbb{N} \cup \{0\}} (A^*)^k \mathbb{Z}^2 = \{0\}$  since

$$\min\{|\xi| : \xi \in (A^*)^k \mathbb{Z}^2 \setminus \{0\}\} \geq cq^k \rightarrow +\infty \text{ as } k \rightarrow +\infty,$$

by Lemma 3.3 applied to  $A^*$ . Therefore,  $T_A$  is a unilateral shift with generating wandering subspace

$$\mathcal{V} = L_0^2(\mathbb{T}^2) \setminus T_A(L_0^2(\mathbb{T}^2)) = \overline{\text{span}}\{e^{2\pi i \xi \cdot x}\}_{\xi \in A^*(\mathbb{Z}^2)}.$$

Hence, Theorem 1.4 applies and, in particular, it applies with  $\varepsilon(n) = (n + 1)^{-\frac{1}{2}}(\log(2 + n))^{-\frac{3}{2} - \eta}$  for any  $\eta > 0$ . The proof now proceeds as in the case of matrices with determinant  $\pm 1$ . Set  $\mathcal{F}_k := A^{*k} \mathbb{Z}^2 \setminus A^{*(k+1)} \mathbb{Z}^2$ . By Lemma (3.3) applied to  $A^*$ , for every positive increasing function  $v$  and  $f$  satisfying (1.4), one has

$$\begin{aligned} \sum_{k \in \mathbb{N} \cup \{0\}} \|\Pi_k f\|_{L^2} &= \sum_{k \in \mathbb{N} \cup \{0\}} \left( \sum_{\xi \in \mathcal{F}_k} |\widehat{f}(\xi)|^2 \right)^{\frac{1}{2}} \\ &\leq \left( \sum_{k \in \mathbb{N} \cup \{0\}} v^{-2}(k) \right)^{\frac{1}{2}} \left( \sum_{k \in \mathbb{N} \cup \{0\}} v^2(k) \sum_{\xi \in \mathcal{F}_k} |\widehat{f}(\xi)|^2 \right)^{\frac{1}{2}} \\ &\leq \left( \sum_{k \in \mathbb{N} \cup \{0\}} v^{-2}(k) \right)^{\frac{1}{2}} \left( \sum_{k \in \mathbb{N} \cup \{0\}} \sum_{\xi \in \mathcal{F}_k} v^2 \left( \frac{\log |\xi| - \log c}{\log q} \right) |\widehat{f}(\xi)|^2 \right)^{\frac{1}{2}}. \end{aligned}$$

The conclusion follows choosing  $v(t) = t^{\frac{1}{2} + \frac{\delta}{2}} + 1$ . In only remains to prove Corollary 1.8, but this is now immediate.

**Proof of Corollary 1.8** Notice that, thanks to (1.5), we can repeat the very same argument of the proof of Theorem 1.7 to obtain the conclusion. ■

### 4 Concluding remarks

As mentioned in the introduction, functions satisfying condition (1.4) on their Fourier coefficients are, for instance, functions in any fractional Sobolev space. Here is another, more general, sufficient condition in terms of the  $L^2$  integral modulus of continuity.

**Proposition 4.1** Let  $\omega(f, t), t > 0$ , be the modulus of continuity of the function  $f \in L^2(\mathbb{T}^d)$ ,

$$\omega(f, t) := \sup_{|y| \leq t} \left( \int_{\mathbb{T}^d} |f(x + y) - f(x)|^2 dx \right)^{\frac{1}{2}}.$$

Also let  $\alpha \geq 0$ . Then there exists a constant  $c$  independent of  $f$  such that

$$\sum_{\xi \in \mathbb{Z}^d} \log^\alpha(1 + |\xi|) |\widehat{f}(\xi)|^2 \leq c \sum_{j=0}^{+\infty} (1 + j^\alpha) \omega^2(f, 2^{-j}).$$



**Proof** One has

$$\sum_{\xi \in \mathbb{Z}^d} \log^\alpha(1 + |\xi|) |\widehat{f}(\xi)|^2 \leq \sum_{j=0}^{+\infty} \log^\alpha(1 + 2^{j+1}) \sum_{2^j \leq |\xi|_\infty < 2^{j+1}} |\widehat{f}(\xi)|^2.$$

It then suffices to show that

$$\sum_{2^j \leq |\xi|_\infty < 2^{j+1}} |\widehat{f}(\xi)|^2 \leq c \omega^2(f, 2^{-j}), \quad \forall j \geq 0.$$

This inequality is well-known, but it is easier to give a proof than a reference. Parseval's identity gives

$$\int_{\mathbb{T}^d} |f(x + y) - f(x)|^2 dx = \sum_{\xi \in \mathbb{Z}^d} |e^{2\pi i \xi y} - 1|^2 |\widehat{f}(\xi)|^2.$$

Write  $\xi = (h, k)$ , with  $h \in \mathbb{Z}$  and  $k \in \mathbb{Z}^{d-1}$ , and take  $y = (2^{-j-2}, 0)$ . Then  $\xi y = 2^{-j-2}h$ , and

$$\sum_{\substack{(h,k) \in \mathbb{Z} \times \mathbb{Z}^d \\ 2^j \leq |h| < 2^{j+1}}} |\widehat{f}(\xi)|^2 \leq c \sum_{\substack{(h,k) \in \mathbb{Z} \times \mathbb{Z}^d \\ 2^j \leq |h| < 2^{j+1}}} |e^{2\pi i 2^{-j-2}h} - 1|^2 |\widehat{f}(\xi)|^2 \leq c \omega^2(f, 2^{-j-2}) \leq c \omega^2(f, 2^{-j}).$$

Iterating for each of the  $d$  coordinates of  $\xi$ , one obtains

$$\sum_{2^j \leq |\xi|_\infty < 2^{j+1}} |\widehat{f}(\xi)|^2 \leq c \omega^2(f, 2^{-j}). \quad \blacksquare$$

It is interesting to observe that, using the above proposition, Corollary 1.8 can be applied whenever  $f$  is the characteristic function of a domain with a fractal boundary with a minimally regular geometry. More precisely, let  $\Omega \subseteq \mathbb{T}^d$  be a Borel measurable set, and suppose that there exists  $\varepsilon > 0$  such that

$$|\{x \in \mathbb{T}^d : \text{dist}(x, \partial\Omega) \leq t\}| \leq c(\log 1/t)^{-2-\varepsilon}, \quad \text{for all } 0 < t \leq 1/2.$$

Notice that this is an assumption on the Minkowski content of  $\partial\Omega$ . Then, for  $0 < \delta < \varepsilon$ ,

$$\begin{aligned} \sum_{j=0}^{+\infty} (1 + j^{1+\delta}) \omega^2(\chi_\Omega, 2^{-j}) &= \sum_{j=0}^{+\infty} (1 + j^{1+\delta}) \sup_{|y| \leq 2^{-j}} \int_{\mathbb{T}^d} |\chi_\Omega(x + y) - \chi_\Omega(x)|^2 dx \\ &\leq \sum_{j=0}^{+\infty} (1 + j^{1+\delta}) |\{x \in \mathbb{T}^d : \text{dist}(x, \partial\Omega) \leq 2^{-j}\}| \\ &\leq c \left( 1 + \sum_{j=1}^{+\infty} (1 + j^{1+\delta}) j^{-2-\varepsilon} \right) < +\infty. \end{aligned}$$

Hence, by Proposition 4.1, we have that

$$\sum_{\xi \in \mathbb{Z}^d} \log^{1+\delta}(1 + |\xi|) |\widehat{\chi_\Omega}(\xi)|^2 < +\infty.$$

Explicitly, from Corollary 1.8, we obtain that, for every matrix  $A$  satisfying (1.5), for every  $\eta > 0$  and for almost every  $x \in \mathbb{T}^d$ , there exists  $C > 0$ , depending on  $x$ ,  $\eta$ , and  $A$ , such that for every  $N \in \mathbb{N}$

$$(4.1) \quad \left| \frac{1}{N} \sum_{n=0}^{N-1} \chi_{\Omega}(A^n x) - |\Omega| \right| \leq CN^{-\frac{1}{2}} \log N^{\frac{3}{2} + \eta}.$$

It is interesting to compare the above estimate with some results in [BCT23]. In particular, in [BCT23, Corollary 4], it is proved that if  $\Omega$  is such that for some  $\beta > 0$  and for every  $t > 0$ , sufficiently small, one has

$$|\{x \in \mathbb{R}^d : \text{dist}(x, \partial\Omega) \leq t\}| \leq ct^{\beta}$$

and if  $\Omega$  satisfies some other mild technical assumptions, then there exists a constant  $c > 0$  such that for every distribution of points  $\{p_n\}_{n=0}^{N-1}$  there exists an affine copy  $\tilde{\Omega}$  of  $\Omega$  which satisfies the estimate

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} \chi_{\tilde{\Omega}}(p_n) - |\tilde{\Omega}| \right| \geq cN^{-\frac{1}{2} - \frac{\beta}{2}}.$$

Moreover, in [BCT23, Theorem 12], it is proved that, under the same hypothesis on  $\Omega$ , there exists  $c > 0$  such that for every  $N$  there exists a distribution of  $N$  points  $\{p_n\}_{n=1}^N$  such that

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} \chi_{\tilde{\Omega}}(p_n) - |\tilde{\Omega}| \right| \leq cN^{-\frac{1}{2} - \frac{\beta}{2}},$$

for “many” affine copies  $\tilde{\Omega}$  of  $\Omega$ . We point out that in this last estimate the single set of points  $\{p_n\}_{n=0}^{N-1}$  depends on  $\Omega$  and  $N$ , while, in our estimate (4.1) the underlying sequence of points is fixed. However, notice that the numerology in our upper bound (4.1) tends to coincide, up to a logarithmic transgression, with the upper bound in [BCT23, Theorem 12] when  $\beta$  tends to 0.

## References

- [AGT10] J. Avigad, P. Gerhardy, and H. Towsner, *Local stability of ergodic averages*. Trans. Amer. Math. Soc. 362(2010), no. 1, 261–288.
- [AI13] J. Avigad and J. Iovino, *Ultraproducts and metastability*. New York J. Math. 19(2013), 713–727.
- [Ais10] C. Aistleitner, *On the law of the iterated logarithm for the discrepancy of lacunary sequences*. Trans. Amer. Math. Soc. 362(2010), no. 11, 5967–5982.
- [Ais13] C. Aistleitner, *On the law of the iterated logarithm for the discrepancy of lacunary sequences II*. Trans. Amer. Math. Soc. 365(2013), no. 7, 3713–3728.
- [AR15] J. Avigad and J. Rute, *Oscillation and the mean ergodic theorem for uniformly convex Banach spaces*. Ergodic Theory Dynam. Systems 35(2015), no. 4, 1009–1027.
- [BAM21] J. Ben-Artzi and B. Morisse, *Uniform convergence in von Neumann’s ergodic theorem in the absence of a spectral gap*. Ergodic Theory Dynam. Systems 41(2021), no. 6, 1601–1611.
- [BBH20] F. Bayart, Z. Buczolich, and Y. Heurteaux, *Fast and slow points of Birkhoff sums*. Ergodic Theory Dynam. Systems 40(2020), no. 12, 3236–3256.
- [BCT23] L. Brandolini, L. Colzani, and G. Travaglini, *Irregularities of distribution for bounded sets and half-spaces*. Mathematika 69(2023), no. 1, 68–89.
- [BHS65] A. Brown, P. R. Halmos, and A. L. Shields, *Cesaro operators*. Acta Sci. Math. 26(1965), 125–137.

- [Bir31] G. D. Birkhoff, *Proof of the ergodic theorem*. Proc. Natl. Acad. Sci. 17(1931), no. 12, 656–660.
- [CGM24] L. Colzani, B. Gariboldi, and A. Monguzzi, *Summability and speed of convergence in an ergodic theorem*. J. Math. Anal. Appl. 536(2024), no. 1, Article no. 128190, 25 pp.
- [Col22] L. Colzani, *Speed of convergence of Weyl sums over Kronecker sequences*. Monatsh. Math. 200(2022), no. 2, 209–228.
- [Cun11] C. Cuny, *Pointwise ergodic theorems with rate with applications to limit theorems for stationary processes*. Stoch. Dyn. 11(2011), no. 1, 135–155.
- [DY18] S. Das and J. A. Yorke, *Super convergence of ergodic averages for quasiperiodic orbits*. Nonlinearity 31(2018), no. 2, 491–501.
- [EW11] M. Einsiedler and T. Ward, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, 259, Springer-Verlag London, Ltd., London, 2011.
- [Fan99] A. H. Fan, *Decay of correlation for expanding toral endomorphisms*, Dynamical Systems. In: L. Wen, and Y. P. Jiang (eds.), Proceedings of the International Conference in Honor of Professor Liao Shantao Peking University, China, World Scientific, 1999.
- [For40] R. Fortet, *Sur une suite également répartie*. Studia Math. 9(1940), 54–70.
- [FS99] A. Furman and Y. Shalom, *Sharp ergodic theorems for group actions and strong ergodicity*. Ergodic Theory Dynam. Systems 19(1999), no. 4, 1037–1061.
- [Kac46] M. Kac, *On the distribution of values of sums of the type  $\sum f(2^k t)$* . Ann. of Math. (2) 47(1946), 33–49.
- [Kat71] Y. Katznelson, *Ergodic automorphisms of  $T^n$  are Bernoulli shifts*. Israel J. Math. 10(1971), 186–195.
- [KP16] A. G. Kachurovskii and I. V. Podvigin, *Estimates of the rate of convergence in the von Neumann and Birkhoff ergodic theorems*. Trans. Moscow Math. Soc. 77(2016), 1–53.
- [KP81] S. Kakutani and K. Petersen, *The speed of convergence in the ergodic theorem*. Monatsh. Math. 91(1981), no. 1, 11–18.
- [Kre79] U. Krengel, *On the speed of convergence in the ergodic theorem*. Monatsh. Math. 86(1978/79), no. 1, 3–6.
- [Krz93] K. Krzyżewski, *On exact toral endomorphisms*. Monatsh. Math. 116(1993), no. 1, 39–47.
- [LA14] T. Löbbecke, *Limit theorems for multivariate lacunary systems*, Preprint, 2014.  
arxiv:1408.2202
- [Mar50] G. Maruyama, *On an asymptotic property of a gap sequence*. Kodai Math. Sem. Rep. 2(1950), 31–32.
- [Mea07] C. Meaney, *Remarks on the Rademacher–Menshov theorem*, CMA/AMSI Research Symposium “Asymptotic Geometric Analysis, Harmonic Analysis, and Related Topics”. In: Proceedings of the Centre for Mathematics and its Applications, 42, Australian National University, Canberra, 2007, pp. 100–110.
- [SNFBK10] B. Sz. Nagy, C. Foias, H. Bercovici, and L. Kérchy, *Harmonic analysis of operators on Hilbert space*. 2nd ed., Springer, New York, 2010.
- [Tao12] T. Tao, *Walsh’s ergodic theorem, metastability, and external Cauchy convergence*. 2012.  
<https://terrytao.wordpress.com/2012/10/25/walshs-ergodic-theorem-metastability-and-external-cauchy-convergence/#more-6236>.
- [vN29] J. von Neumann, *Zur Theorie der unbeschränkten Matrizen*. J. Reine Angew. Math. 161(1929), 208–236.
- [vN32] J. von Neumann, *Proof of the quasi-ergodic hypothesis*. Proc. Natl. Acad. Sci. 18(1932), no. 1, 70–82.
- [Zyg03] A. Zygmund, *Trigonometric series*. 3rd ed., Cambridge University Press, 2003.

Dipartimento di Matematica e Applicazioni, Università degli Studi di Milano–Bicocca, Milano, Italy  
e-mail: nikolaos.chalmoukis@unimib.it leonardo.colzani@unimib.it

Dipartimento di Ingegneria Gestionale, dell’Informazione e della Produzione, Univesità degli Studi di Bergamo, Dalmine (BG), Italy  
e-mail: biancamaria.gariboldi@unibg.it alessandro.monguzzi@unibg.it