Rev. Mat. Iberoam. 38 (2022), no. 7, 2249–2284
DOI 10.4171/RMI/1383

Oscillation inequalities in ergodic theory and analysis: one-parameter and multi-parameter perspectives

Mariusz Mirek, Tomasz Z. Szarek and James Wright

Abstract. In this survey we review useful tools that naturally arise in the study of pointwise convergence problems in analysis, ergodic theory and probability. We will pay special attention to quantitative aspects of pointwise convergence phenomena from the point of view of oscillation estimates in both the single and several parameter settings. We establish a number of new oscillation inequalities and give new proofs for known results with elementary arguments.

In honour of Antonio Córdoba and José Luis Fernández.

1. Introduction

Pointwise convergence is the most natural as well as the most difficult type of convergence to establish. It requires sophisticated tools in analysis, ergodic theory and probability. In this survey, we will review variation and oscillation semi-norms as well as the λ-jump counting function which give us quantitative measures for pointwise convergence. However, we will concentrate on the central role that oscillation inequalities play, both in the one-parameter and multi-parameter settings.

In the one-parameter setting, we derive a simple abstract oscillation estimate for the so-called projective operators, which will result in oscillation estimates for martingales, smooth bump functions as well as the Carleson operator. The multi-parameter oscillation semi-norm is the only available tool that allows us to handle efficiently multi-parameter pointwise convergence problems with arithmetic features. This contrasts sharply with the one-parameter setting, where we have a variety of tools including oscillations, variations or λ-jumps to handle pointwise convergence problems. The multi-parameter oscillation estimates will be illustrated in the context of the Dunford–Zygmund ergodic theorem for commuting measure-preserving transformations, as well as observations of Bourgain for certain multi-parameter polynomial ergodic averages.

2020 Mathematics Subject Classification: Primary 37A30; Secondary 37A46, 42B25.
Keywords: Ergodic average, (pointwise) ergodic theorem, maximal, variational, jump, oscillation estimate.

We begin with describing methods that permit us to handle pointwise convergence problems in the context of various ergodic averaging operators. Before we do this, we set up notation and terminology, which will allow us to discuss various concepts in a fairly unified way.

Throughout this survey, the triple $(X, \mathcal{B}(X), \mu)$ denotes a σ-finite measure space. The space of all formal k-variate polynomials $P(\mathrm{m}_1, \dots, \mathrm{m}_k)$ with $k \in \mathbb{Z}_+$ indeterminates $\mathrm{m}_1, \dots, \mathrm{m}_k$ and integer coefficients will be denoted by $\mathbb{Z}[\mathrm{m}_1, \dots, \mathrm{m}_k]$. We will always identify each polynomial $P \in \mathbb{Z}[\mathrm{m}_1, \dots, \mathrm{m}_k]$ with a function $(m_1, \dots, m_k) \mapsto P(m_1, \dots, m_k)$ from $\mathbb{Z}^k$ to $\mathbb{Z}$.

Let $d, k \in \mathbb{Z}_+$. Consider a family $\mathcal{T} = (T_1, \dots, T_d)$ of invertible commuting measure-preserving transformations on X, polynomials $\mathcal{P} = (P_1, \dots, P_d) \subset \mathbb{Z}[\mathrm{m}_1, \dots, \mathrm{m}_k]$, an integer k-tuple $M = (M_1, \dots, M_k) \in \mathbb{Z}_+^k$, and a measurable function $f : X \to \mathbb{C}$. We consider the multi-parameter polynomial ergodic average

$$\frac{1}{M_1 \cdots M_k} \sum_{m_1=1}^{M_1} \cdots \sum_{m_k=1}^{M_k} f\big(T_1^{P_1(m_1,\dots,m_k)} \cdots T_d^{P_d(m_1,\dots,m_k)} x\big).$$

We denote this average by $A^{\mathcal{P}}_{M;X,\mathcal{T}} f(x)$, and we use the notation

$$A^{\mathcal{P}}_{M;X,\mathcal{T}} f(x) := \mathbb{E}_{m \in Q_M} f(T_1^{P_1(m)} \cdots T_d^{P_d(m)} x), \quad x \in X, \tag{1.1}$$

where $Q_M := [M_1] \times \cdots \times [M_k]$ is a box in $\mathbb{Z}^k$ with $[N] := (0, N] \cap \mathbb{Z}$, for any real number $N \geq 1$, and $\mathbb{E}_{y \in Y} f(y) := \frac{1}{\#Y} \sum_{y \in Y} f(y)$ for any finite set Y and any $f : Y \to \mathbb{C}$. We will often abbreviate $A^{\mathcal{P}}_{M;X,\mathcal{T}}$ to $A^{\mathcal{P}}_{M;X}$ when the transformations are understood. Depending on how explicit we want to be, more precision may be necessary and we will write out the averages

$$A^{\mathcal{P}}_{M;X} f(x) = A^{P_1,\dots,P_d}_{M_1,\dots,M_k;X} f(x) \quad \text{or} \quad A^{\mathcal{P}}_{M;X,\mathcal{T}} f(x) = A^{P_1,\dots,P_d}_{M_1,\dots,M_k;X,T_1,\dots,T_d} f(x).$$

Example 1.2. Due to the Calderón transference principle [13], the most important dynamical system, from the point of view of pointwise convergence problems, is the integer shift system. Namely, it is the d-dimensional lattice $(\mathbb{Z}^d, \mathcal{B}(\mathbb{Z}^d), \mu_{\mathbb{Z}^d})$ equipped with a family of shifts $S_1, \dots, S_d : \mathbb{Z}^d \to \mathbb{Z}^d$, where $\mathcal{B}(\mathbb{Z}^d)$ denotes the σ-algebra of all subsets of $\mathbb{Z}^d$, $\mu_{\mathbb{Z}^d}$ denotes counting measure on $\mathbb{Z}^d$, and $S_j(x) := x - e_j$ for every $x \in \mathbb{Z}^d$ (here e_j is the jth basis vector from the standard basis in $\mathbb{Z}^d$, for each $j \in [d]$). Then the average $A^{\mathcal{P}}_{M;X,\mathcal{T}}$ from (1.1) with $\mathcal{T} = (T_1, \dots, T_d) = (S_1, \dots, S_d)$ can be rewritten for any $x = (x_1, \dots, x_d) \in \mathbb{Z}^d$ and any finitely supported function $f : \mathbb{Z}^d \to \mathbb{C}$ as

$$A^{\mathcal{P}}_{M;\mathbb{Z}^d} f(x) = \mathbb{E}_{m \in Q_M} f(x_1 - P_1(m), \dots, x_d - P_d(m)). \tag{1.3}$$

1.1. Birkhoff's and von Neumann's ergodic theorems

In the early 1930's, Birkhoff [5] and von Neumann [61] established an almost everywhere pointwise ergodic theorem and a mean ergodic theorem, respectively, which we summarize in the following result.

Theorem 1.4 (Birkhoff's and von Neumann's ergodic theorem). *Let $(X, \mathcal{B}(X), \mu)$ be a σ-finite measure space equipped with a measure-preserving transformation $T: X \to X$. Then for every $p \in [1, \infty)$ and every $f \in L^p(X)$, the averages*

$$A^{\mathrm{m}}_{M;X,T} f(x) = \mathbb{E}_{m \in [M]} f(T^m x), \quad x \in X,\ M \in \mathbb{Z}_+,$$

converge almost everywhere on X and in $L^p(X)$ norm as $M \to \infty$.

Although there are many proofs of Theorem 1.4 in the literature (we refer for instance to the monographs [21, 64] for more details and the historical background), there is a particular proof which is important in our context. This proof illustrates the classical strategy for handling pointwise convergence problems, which is based on a two-step procedure:

(i) The first step establishes $L^p(X)$ boundedness (for $p \in (1, \infty)$), or a weak type $(1, 1)$ bound (when $p = 1$) of the corresponding maximal function $\sup_{M \in \mathbb{Z}_+} |A^{\mathrm{m}}_{M;X,T} f(x)|$. This in turn, using the Calderón transference principle [13], can be derived from the corresponding maximal bounds for $\sup_{M \in \mathbb{Z}_+} |A^{\mathrm{m}}_{M;\mathbb{Z}} f(x)|$, the Hardy–Littlewood maximal function on the set of integers, see (1.3). Having these maximal estimates in hand, one can easily prove that the set

$$\mathbf{PC}[L^p(X)] = \big\{ f \in L^p(X) : \lim_{M \to \infty} A^{\mathrm{m}}_{M;X,T} f \text{ exists } \mu\text{-almost everywhere on } X \big\}$$

is closed in $L^p(X)$.

(ii) In the second step, one shows that $\mathbf{PC}[L^p(X)] = L^p(X)$. In view of the first step, the task is reduced to finding a dense class of functions in $L^p(X)$ for which we have pointwise convergence. In our problem, let us first assume $p = 2$. Then invoking a variant of the Riesz decomposition [69], a good candidate is the space $\mathrm{I}_T \oplus \mathrm{J}_T \subseteq L^2(X)$, where

$$\begin{aligned} \mathrm{I}_T &:= \{ f \in L^2(X) : f \circ T = f \}, \\ \mathrm{J}_T &:= \{ g - g \circ T : g \in L^2(X) \cap L^\infty(X) \}. \end{aligned}$$

We then note that $A^{\mathrm{m}}_{M;\mathbb{Z}} f = f$ for $f \in \mathrm{I}_T$, and $\lim_{M \to \infty} A^{\mathrm{m}}_{M;X,T} h = 0$ for $h \in \mathrm{J}_T$, since

$$A^{\mathrm{m}}_{M;X,T} h = M^{-1}(g \circ T - g \circ T^{M+1})$$

telescopes, whenever $h = g - g \circ T \in \mathrm{J}_T$. This establishes pointwise almost everywhere convergence of $A^{\mathrm{m}}_{M;X,T}$ on $\mathrm{I}_T \oplus \mathrm{J}_T$, which is dense in $L^2(X)$. These two steps guarantee that $\mathbf{PC}[L^2(X)] = L^2(X)$. Consequently, $A^{\mathrm{m}}_{M;X,T}$ converges pointwise on $L^p(X) \cap L^2(X)$ for any $p \in [1, \infty)$. Since $L^p(X) \cap L^2(X)$ is dense in $L^p(X)$, we also conclude, in view of the first step, that $\mathbf{PC}[L^p(X)] = L^p(X)$, and this completes a brief outline of the proof of Theorem 1.4.

1.2. Dunford–Zygmund pointwise ergodic theorem

In the early 1950's, it was observed by Dunford [19] and independently by Zygmund [77] that the two-step procedure can be applied in a multi-parameter setting. More precisely, the Dunford–Zygmund multi-parameter pointwise ergodic theorem, where the convergence is understood in the unrestricted sense, can be formulated as follows.

Theorem 1.5 (Dunford–Zygmund ergodic theorem). *Let $d \in \mathbb{Z}_+$ and let $(X, \mathcal{B}(X), \mu)$ be a σ-finite measure space equipped with a family $\mathcal{T} = (T_1, \ldots, T_d)$ of not necessarily commuting and measure-preserving transformations $T_1, \ldots, T_d : X \to X$. Then for every $p \in (1, \infty)$ and every $f \in L^p(X)$, the averages*

$$A^{\mathrm{m}_1,\ldots,\mathrm{m}_d}_{M_1,\ldots,M_d;X,\mathcal{T}} f(x) = \mathbb{E}_{m \in Q_M} f(T_1^{m_1} \cdots T_d^{m_d} x), \quad x \in X,\ M = (M_1, \ldots, M_d) \in \mathbb{Z}_+^d,$$

converge almost everywhere on X and in $L^p(X)$ norm as $\min\{M_1, \ldots, M_d\} \to \infty$.

This theorem has a fairly simple proof, which is based on the following identity:

$$A^{\mathrm{m}_1,\ldots,\mathrm{m}_d}_{M_1,\ldots,M_d;X,\mathcal{T}} f = A^{\mathrm{m}_1}_{M_1;X,T_1} \circ \cdots \circ A^{\mathrm{m}_d}_{M_d;X,T_d} f.$$

The $L^p(X)$ bounds for the strong maximal function $\sup_{M \in \mathbb{Z}_+^d} |A^{\mathrm{m}_1,\ldots,\mathrm{m}_d}_{M_1,\ldots,M_d;X,\mathcal{T}} f|$, for $p \in (1, \infty]$, follow easily by applying d times the corresponding $L^p(X)$ bounds for $\sup_{M \in \mathbb{Z}_+} |A^{\mathrm{m}}_{M;X,T} f|$. This establishes the first step in the two-step procedure described above. The second step is based on a suitable adaptation of the telescoping argument to the multi-parameter setting and an application of the classical Birkhoff ergodic theorem, see [62] for more details. These two steps establish Theorem 1.5 and motivate our further discussion on multi-parameter convergence problems. One also knows that pointwise convergence in Theorem 1.5 may fail if $p = 1$, and that the operator $f \mapsto \sup_{M \in \mathbb{Z}_+^d} |A^{\mathrm{m}_1,\ldots,\mathrm{m}_d}_{M_1,\ldots,M_d;X,\mathcal{T}} f|$ is not of weak type $(1, 1)$ in general (even if we assume that the transformations T_j, $1 \le j \le d$, commute). A model example is $X = \mathbb{Z}^d$ and $T_j x = x - e_j$, $1 \le j \le d$, where e_j is the jth coordinate vector. Then the corresponding maximal operator is just the strong maximal operator, for which it is well known that the weak type $(1, 1)$ estimate does not hold.

1.3. Quantitative tools in the study of pointwise convergence

The approach described in the context of Theorem 1.4 and Theorem 1.5 has a quantitative nature, but it says nothing quantitatively about pointwise convergence. This approach is very effective in pointwise convergence questions arising in harmonic analysis, as there are many natural dense subspaces in Euclidean settings which can be used to establish pointwise convergence. However, for ergodic theoretic questions, when one works with abstract measure spaces, the situation is dramatically different, as Bourgain showed [6–8]. We shall see more examples below.

Consequently, the second step from the two-step procedure may require more quantitative tools to establish pointwise convergence. To overcome the difficulties with determining dense subspace for which pointwise convergence may be verified, Bourgain [8] proposed three other approaches.

(1) The first approach is based on controlling the so-called oscillation semi-norms. Let $\mathbb{J} \subseteq \mathbb{N}$ be so that $\#\mathbb{J} \ge 2$, let $I = (I_j : j \in \mathbb{N}_{\le J})$ be a strictly increasing sequence of length $J + 1$ for some $J \in \mathbb{Z}_+$, which takes values in $\mathbb{J}$, and recall that for any sequence $(\mathfrak{a}_t : t \in \mathbb{J}) \subseteq \mathbb{C}$, and any exponent $1 \le r < \infty$, the r-oscillation seminorm is defined by

$$O^r_{I,J}(\mathfrak{a}_t : t \in \mathbb{J}) := \Big(\sum_{j=0}^{J-1} \sup_{\substack{I_j \le t < I_{j+1} \\ t \in \mathbb{J}}} |\mathfrak{a}_t - \mathfrak{a}_{I_j}|^r \Big)^{1/r}. \tag{1.6}$$

We will give a more general definition of r-oscillations in the multi-parameter setting; see (2.3).

(2) The second approach is based on controlling the so-called r-variation seminorms. For any $\mathbb{I} \subseteq \mathbb{N}$, any sequence $(\mathfrak{a}_t : t \in \mathbb{I}) \subseteq \mathbb{C}$, and any exponent $1 \le r < \infty$, the r-variation semi-norm is defined to be

$$V^r(\mathfrak{a}_t : t \in \mathbb{I}) := \sup_{J \in \mathbb{Z}_+} \sup_{\substack{t_0 < \cdots < t_J \\ t_j \in \mathbb{I}}} \Big(\sum_{j=0}^{J-1} |\mathfrak{a}_{t_{j+1}} - \mathfrak{a}_{t_j}|^r \Big)^{1/r},$$

where the latter supremum is taken over all finite increasing sequences in $\mathbb{I}$.

(3) The third approach is based on studying the λ-jump counting function, which is closely related to r-variations. For any $\mathbb{I} \subseteq \mathbb{N}$ and any $\lambda > 0$, the λ-jump counting function of a sequence $(\mathfrak{a}_t : t \in \mathbb{I}) \subseteq \mathbb{C}$ is defined by

$$N_\lambda(\mathfrak{a}_t : t \in \mathbb{I}) := \sup\big\{ J \in \mathbb{N} : \exists_{\substack{t_0 < \cdots < t_J \\ t_j \in \mathbb{I}}} : \min_{0 \le j \le J-1} |\mathfrak{a}_{t_{j+1}} - \mathfrak{a}_{t_j}| \ge \lambda \big\}.$$

We also refer to Section 2 for simple properties of r-oscillations, r-variations and λ-jumps. These will be illustrated in the context of bounded martingales, a toy model explaining their quantitative nature and their usefulness in pointwise convergence problems.

1.4. Bourgain's pointwise ergodic theorem

In the early 1980's, Bellow [1] (being motivated by some problems from equidistribution theory), and independently Furstenberg [23] (being motivated by some problems from additive combinatorics in the spirit of Szemerédi's theorem [73] for arithmetic progressions), posed the problem of whether for any polynomial $P \in \mathbb{Z}[\mathrm{m}]$ and any measure-preserving map $T : X \to X$ on a probability space $(X, \mathcal{B}(X), \mu)$, the averages

$$A^{P(\mathrm{m})}_{M;X,T} f(x) = \mathbb{E}_{m \in [M]} f(T^{P(m)} x), \quad x \in X,\ M \in \mathbb{Z}_+, \tag{1.7}$$

converge almost everywhere on X as $M \to \infty$, for any $f \in L^\infty(X)$.

An affirmative answer to this question was given by Bourgain in a series of groundbreaking papers [6–8] which we summarize in the following theorem.

Theorem 1.8 (Bourgain's ergodic theorem). *Let $(X, \mathcal{B}(X), \mu)$ be a σ-finite measure space equipped with an invertible measure-preserving transformation $T : X \to X$. Assume that $P \in \mathbb{Z}[\mathrm{m}]$ is a polynomial such that $P(0) = 0$. Then for every $p \in (1, \infty)$ and every $f \in L^p(X)$, the averages $A^P_{M;X,T} f$ from (1.7) converge almost everywhere on X and in $L^p(X)$ norm as $M \to \infty$.*

Theorem 1.8 is an instance where establishing pointwise convergence on a dense class is a challenging problem. The decomposition $\mathrm{I}_T \oplus \mathrm{J}_T$ of von Neumann (as for $A^{\mathrm{m}}_{M;X,T}$) is not sufficient if $\deg P \ge 2$, though it still makes sense. Even for the squares $P(m) = m^2$, it is not clear whether $\lim_{M\to\infty} A^{\mathrm{m}^2}_{M;X,T} h = 0$ for $h \in \mathrm{J}_T$. The reason is that the averages $A^{\mathrm{m}^2}_{M;X,T} h$ do not telescope for $h \in \mathrm{J}_T$ anymore, since the differences $(m+1)^2 - m^2 = 2m + 1$ have unbounded gaps.

Nearly two decades after Bourgain papers [6–8], it was discovered that the range of $p \in (1, \infty)$ in Bourgain's theorem is sharp. In contrast to Birkhoff's theorem, if $P \in \mathbb{Z}[\mathrm{m}]$ is a polynomial of degree at least two, the pointwise convergence at the endpoint for $p = 1$ may fail as was shown by Buczolich and Mauldin [10] for $P(m) = m^2$ and by LaVictoire [49] for $P(m) = m^k$ for any $k \geq 2$. This also stands in sharp contrast to what happens for continuous analogues of ergodic averages, and shows that any intuition that we build in Euclidean harmonic analysis (when sums are replaced with integrals) can fail dramatically in discrete problems.

Bourgain [6–8] also used the two-step procedure to prove Theorem 1.8. In the first step, it was proved that for all $p \in (1, \infty]$, there exists $C_{p,P} > 0$ such that for every $f \in L^p(X)$ we have

$$\Big\| \sup_{M \in \mathbb{Z}_+} |A^P_{M;X,T} f| \Big\|_{L^p(X)} \leq C_{p,P} \|f\|_{L^p(X)}. \tag{1.9}$$

However, in the second step of the two-step procedure, a quantitative pointwise ergodic theorem was established by studying oscillation semi-norms, see (1.6). More, precisely, it was proved that for any $\tau > 1$, any sequence of integers $I = (I_j : j \in \mathbb{N}) \subseteq \mathbb{L}_\tau := \{\lfloor \tau^n \rfloor : n \in \mathbb{N}\}$ such that $I_{j+1} > 2I_j$ for all $j \in \mathbb{N}$, and any $f \in L^2(X)$, one has

$$\|O^2_{I,J}(A^P_{M;X,T} f : M \in \mathbb{L}_\tau)\|_{L^2(X)} \leq C_{I,\tau}(J) \|f\|_{L^2(X)}, \quad J \in \mathbb{Z}_+, \tag{1.10}$$

where $C_{I,\tau}(J)$ is a constant depending on I and τ that satisfies

$$\lim_{J \to \infty} J^{-1/2} C_{I,\tau}(J) = 0. \tag{1.11}$$

Bourgain [6–8] had the ingenious insight to see that inequality (1.10) with (1.11) suffices to establish pointwise convergence of $A^P_{M;X,T} f$ for any $f \in L^2(X)$. Inequality (1.10) with (1.11) can be thought of as the weakest possible quantitative form for pointwise convergence. On the one hand, (1.10) is very close to the maximal inequality, since by using (1.9) with $p = 2$ we can derive (1.10) with a constant at most $J^{1/2}$. On the other hand, any improvement (better than $J^{1/2}$) for the constant in (1.10) implies (1.11) and so ensures pointwise convergence of $A^P_{M;X,T} f$ for any $f \in L^2(X)$, see Proposition 2.8, where the details, even in the multi-parameter setting, are given. Therefore, from this point of view, inequality (1.10) with (1.11) is the minimal quantitative requirement necessary to establish pointwise convergence.

Bourgain's papers [6–8] were a significant breakthrough in ergodic theory, which used a variety of new tools (ranging from harmonic analysis and number theory through probability and the theory of Banach spaces) to study pointwise convergence problems in analysis understood in a broad sense. In [8], a complete proof of Theorem 1.8 is given using the notions of r-variations and λ-jumps (introduced by Pisier and Xu [66]), which are two important quantitative tools in the study of pointwise convergence problems. This initiated a systematic study of quantitative estimates in harmonic analysis and ergodic theory which resulted in a vast literature: in ergodic theory [10,33–36,42,43,49,54,55,59], in discrete harmonic analysis [29–32,51,53,57,60,65,70], and in classical harmonic analysis [2, 3, 14, 17, 26, 37, 38, 45, 46, 56, 58, 63].

Not long after [8], Lacey refined Bourgain's argument (see Theorem 4.23 on p. 95 of [70]), and showed that for every $\tau > 1$ there is a constant $C_\tau > 0$ such that for any $f \in L^2(X)$ one has

$$\sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{L}_\tau)} \|O^2_{I,J}(A^P_{M;X,T}f : M \in \mathbb{L}_\tau)\|_{L^2(X)} \le C_\tau \|f\|_{L^2(X)}, \tag{1.12}$$

where $\mathfrak{S}_J(\mathbb{L}_\tau)$ denotes the set of all strictly increasing sequences $I = (I_j : j \in \mathbb{N}_{\le J}) \subset \mathbb{L}_\tau$ of length $J + 1$ for some $J \in \mathbb{Z}_+$. Inequality (1.12) was the first uniform oscillation result in the class of τ-lacunary sequences. Lacey's observation naturally motivated a question (which also motivates this survey) whether there are uniform estimates, independent of $\tau > 1$, of oscillation inequalities in (1.12). For the Birkhoff averages $A^{\mathrm{m}}_{M;X,T}$, this was explicitly formulated in Problem 4.12 on p. 80 of [70]. We will discuss below uniform oscillation estimates as well as other quantitative forms of pointwise convergence including r-variations and λ-jumps.

1.5. Martingales: a model to study pointwise convergence problems

In order to understand the relationship between r-oscillations, r-variations and λ-jumps, we will use bounded martingales $\mathfrak{f} = (\mathfrak{f}_n : X \to \mathbb{C} : n \in \mathbb{Z}_+)$ as a toy model to help us understand the connections and various nuances. All properties that will be used in the discussion below are collected in Section 2. The discussion will follow the development of the various notions in chronological order.

The r-variations for $\mathfrak{f} = (\mathfrak{f}_n : X \to \mathbb{C} : n \in \mathbb{Z}_+)$ were investigated by Lépingle [50], who established that for all $r \in (2,\infty)$ and $p \in (1,\infty)$, there is a constant $C_{p,r} > 0$ such that

$$\|V^r(\mathfrak{f}_n : n \in \mathbb{Z}_+)\|_{L^p(X)} \le C_{p,r} \sup_{n\in\mathbb{Z}_+} \|\mathfrak{f}_n\|_{L^p(X)}. \tag{1.13}$$

In fact, Lépingle [50] also proved a weak type $(1,1)$ estimate. A counterexample from [38] for $r = 2$ shows that (1.13) holds with sharp ranges of exponents. This counterexample plays an important role showing that r-variation estimates only hold when $r > 2$. In fact, this is the best we can expect in applications in analysis and ergodic theory.

Inequality (1.13) can be thought of as an extension of Doob's maximal inequality for martingales, which gives a quantitative form of the martingale convergence theorem. Indeed, on the one hand, inequality (1.13) implies that the sequences $(\mathfrak{f}_n : n \in \mathbb{Z}_+)$ converges almost everywhere on X as $n \to \infty$. On the other hand, one has

$$\Big\| \sup_{n\in\mathbb{Z}_+} |\mathfrak{f}_n| \Big\|_{L^p(X)} \le \|V^r(\mathfrak{f}_n : n \in \mathbb{Z}_+)\|_{L^p(X)} + \|\mathfrak{f}_{n_0}\|_{L^p(X)}$$

for any $n_0 \in \mathbb{Z}_+$ (see (2.14) below), which shows that r-variational estimates lie deeper than maximal function estimates. We refer to [8, 57, 66] for generalizations and different proofs of (1.13).

Interestingly, Bourgain [8] gave a new proof of inequality (1.13), where it was used to address the issue of pointwise convergence of $A^P_{M;X,T}f$, see (1.7). This initiated a systematic study of r-variations and other quantitative estimates in harmonic analysis and ergodic theory, which resulted in a vast literature [33, 34, 37, 38, 55, 57–59, 63, 76], and recently [29, 43, 54]. Due to (2.19) below, one has

$$\sup_{\lambda>0} \|\lambda N_\lambda(\mathfrak{f}_n : n \in \mathbb{Z}_+)^{1/r}\|_{L^p(X)} \le \|V^r(\mathfrak{f}_n : n \in \mathbb{Z}_+)\|_{L^p(X)}, \tag{1.14}$$

which combined with (1.13) implies λ-jump inequalities for martingales for any $r > 2$. Although the right-hand side of (1.14) blows up when $r \to 2$, it is possible to prove that for every $p \in (1, \infty)$, there exists a constant $C_p > 0$ such that

$$\sup_{\lambda>0} \|\lambda N_\lambda(\mathfrak{f}_n : n \in \mathbb{Z}_+)^{1/2}\|_{L^p(X)} \le C_p \sup_{n\in\mathbb{Z}_+} \|\mathfrak{f}_n\|_{L^p(X)}. \tag{1.15}$$

Inequality (1.15) was first established by Pisier and Xu [66] on $L^2(X)$, and then extended by Bourgain (inequality (3.5) in [8]) on $L^p(X)$ for all $p \in (1, \infty)$. In fact, Bourgain used (1.15) to prove (1.13) by noting that (1.14) can be reversed in the sense that for every $p \in [1, \infty]$ and $1 \le \rho < r \le \infty$, one has

$$\|V^r(\mathfrak{f}_n : n \in \mathbb{Z}_+)\|_{L^{p,\infty}(X)} \lesssim_{p,\rho,r} \sup_{\lambda>0} \|\lambda N_\lambda(\mathfrak{f}_n : n \in \mathbb{Z}_+)^{1/\rho}\|_{L^{p,\infty}(X)}, \tag{1.16}$$

which follows from (2.20) below. One cannot replace $L^{p,\infty}(X)$ with $L^p(X)$ in (1.16), see [54] for more details. Combining (1.15) and (1.16) with $\rho = 2$ and interpolating, one obtains (1.13). Therefore uniform λ-jump estimates from (1.15) can be thought of as end-point estimates for r-variations where we have seen that r-variations may be unbounded at the endpoint in question. We have already noted the failure of Lépingle's inequality (1.13) when $r = 2$.

Even though we have a fairly complete picture of the relationship between r-variations and λ-jumps, the relations with r-oscillations are less obvious. It follows from (2.15) below that

$$\sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{Z}_+)} \|O^r_{I,J}(\mathfrak{f}_n : n \in \mathbb{Z}_+)\|_{L^p(X)} \le \|V^r(\mathfrak{f}_n : n \in \mathbb{Z}_+)\|_{L^p(X)}, \tag{1.17}$$

where $\mathfrak{S}_J(\mathbb{Z}_+)$ denotes the set of all strictly increasing sequences $I = (I_j : j \in \mathbb{N}_{\le J}) \subset \mathbb{Z}_+$ of length $J+1$ for some $J \in \mathbb{Z}_+$. In view of (1.13), this immediately implies r-oscillations estimates for martingales on $L^p(X)$ for all $r \in (2, \infty)$ and $p \in (1, \infty)$.

It was shown by Jones, Kaufman, Rosenblatt and Wierdl (Theorem 6.4 on p. 930 of [33]) that for every $p \in (1, \infty)$ there is a constant $C_p > 0$ such that

$$\sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{Z}_+)} \|O^2_{I,J}(\mathfrak{f}_n : n \in \mathbb{Z}_+)\|_{L^p(X)} \le C_p \sup_{n\in\mathbb{Z}_+} \|\mathfrak{f}_n\|_{L^p(X)}. \tag{1.18}$$

Inequality (1.18) is also an extension of Doob's maximal inequality for martingales, as one has

$$\Big\| \sup_{n\in\mathbb{Z}_+} |\mathfrak{f}_n| \Big\|_{L^p(X)} \le \sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{Z}_+)} \|O^2_{I,J}(\mathfrak{f}_n : n \in \mathbb{Z}_+)\|_{L^p(X)} + \|\mathfrak{f}_{n_0}\|_{L^p(X)}$$

for any $n_0 \in \mathbb{Z}_+$. This follows from Proposition 2.6 below. Moreover, in view of Proposition 2.8, inequality (1.18) also gives a quantitative form of the martingale convergence theorem.

In Section 3 we give a new proof of inequality (1.18), which follows from an abstract result formulated for certain projections, see Theorem 3.1 in Section 3. This abstract theorem will also establish oscillation inequalities for smooth bump functions (see Proposition 3.13 and Theorem 3.17), and establish oscillation inequalities for the Carleson operator (see Proposition 3.34 as well as Proposition 3.22). It will also show that oscillation estimates are very close to maximal estimates even though it follows from Proposition 2.6 that oscillations always dominate maximal functions, see the discussion below Theorem 1.8.

Inequalities (1.17) and (1.18) are similar to inequalities (1.14) and (1.15), respectively, and this raises a natural question whether 2-oscillations can be interpreted as an endpoint for r-variations when $r > 2$ in the sense of inequality (1.16). Recently this problem was investigated in [54, Theorem 1.9] and answered in the negative. Specifically, one can show if $1 \le p < \infty$ and $1 < \rho \le r < \infty$ are fixed, then it is *not true* that the estimates

$$\begin{aligned} &\sup_{\lambda>0} \|\lambda N_\lambda(f(\cdot,t) : t \in \mathbb{N})^{1/r}\|_{\ell^{p,\infty}(\mathbb{Z})} \le C_{p,\rho,r} \sup_{I\in\mathfrak{S}_\infty(\mathbb{N})} \|O^\rho_{I,\infty}(f(\cdot,t) : t \in \mathbb{N})\|_{\ell^p(\mathbb{Z})}, \\ &\|V^r(f(\cdot,t) : t \in \mathbb{N})\|_{\ell^{p,\infty}(\mathbb{Z})} \le C_{p,\rho,r} \sup_{I\in\mathfrak{S}_\infty(\mathbb{N})} \|O^\rho_{I,\infty}(f(\cdot,t) : t \in \mathbb{N})\|_{\ell^p(\mathbb{Z})} \end{aligned} \tag{1.19}$$

hold uniformly for every measurable function $f : \mathbb{Z} \times \mathbb{N} \to \mathbb{R}$. The failure of the inequalities (1.19) shows that the space induced by ρ-oscillations is different from the spaces induced by r-variations and λ jumps whenever $\rho \le r$. Also, the failure of the inequalities (1.19) shows that ρ-oscillation inequalities cannot be seen (at least in a straightforward way, understood in the sense of inequality (1.16)) as endpoint estimates for r-variations, though it still makes sense to ask whether a priori bounds for 2-oscillations imply bounds for r-variations for any $r > 2$. This is an intriguing question from the point of view of quantitative pointwise convergence problems. If true, it would reduce pointwise convergence problems to the study of 2-oscillations, which in certain cases are simpler since they are closer to square functions.

1.6. Quantitative forms of Bourgain's ergodic theorem

Quantitative bounds in the context of ergodic polynomial averaging operators have been intensively studied over the last decade. These investigations were the subject of the following papers [54–56, 59], which generalized Bourgain's papers [6–8] in various ways, and can be summarized as follows.

Theorem 1.20. *Let $d, k \in \mathbb{Z}_+$ and $\mathcal{P} = (P_1, \dots, P_d) \subset \mathbb{Z}[\mathrm{m}_1, \dots, \mathrm{m}_k]$ such that $P_j(0) = 0$ for $j \in [d]$ be given. Let $(X, \mathcal{B}(X), \mu)$ be a σ-finite measure space endowed with a family $\mathcal{T} = (T_1, \dots, T_d)$ of commuting invertible measure-preserving transformations on X. Let $f \in L^p(X)$ for some $1 \le p \le \infty$, and for $M \in \mathbb{Z}_+$, let $A^{\mathcal{P}}_{M;X,\mathcal{T}} f = A^{P_1,\dots,P_d}_{M_1,\dots,M_k;X,T_1,\dots,T_d} f$ be the polynomial ergodic average defined in (1.1) with parameters $M_1 = \dots = M_k = M$.*

- (i) (Mean ergodic theorem) *If $1 < p < \infty$, then the averages $A^{\mathcal{P}}_{M;X,\mathcal{T}} f$ converge in $L^p(X)$ norm as $M \to \infty$.*
- (ii) (Pointwise ergodic theorem) *If $1 < p < \infty$, then the averages $A^{\mathcal{P}}_{M;X,\mathcal{T}} f$ converge pointwise almost everywhere on X as $M \to \infty$.*

(iii) (Maximal ergodic theorem) *If* $1 < p \le \infty$, *then one has*

$$\Big\| \sup_{M\in\mathbb{Z}_+} |A^{\mathcal{P}}_{M;X,\mathcal{T}} f| \Big\|_{L^p(X)} \lesssim_{d,k,p,\deg \mathcal{P}} \|f\|_{L^p(X)}. \tag{1.21}$$

(iv) (Variational ergodic theorem) *If* $1 < p < \infty$ *and* $2 < r < \infty$, *then one has*

$$\|V^r(A^{\mathcal{P}}_{M;X,\mathcal{T}} f : M \in \mathbb{Z}_+)\|_{L^p(X)} \lesssim_{d,k,p,r,\deg \mathcal{P}} \|f\|_{L^p(X)}. \tag{1.22}$$

(v) (Jump ergodic theorem) *If* $1 < p < \infty$, *then one has*

$$\sup_{\lambda>0} \|\lambda N_\lambda(A^{\mathcal{P}}_{M;X,\mathcal{T}} f : M \in \mathbb{Z}_+)^{1/2}\|_{L^p(X)} \lesssim_{d,k,p,\deg \mathcal{P}} \|f\|_{L^p(X)}. \tag{1.23}$$

(vi) (Oscillation ergodic theorem) *If* $1 < p < \infty$, *then one has*

$$\sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{Z}_+)} \|O^2_{I,J}(A^{\mathcal{P}}_{M;X,\mathcal{T}} f : M \in \mathbb{Z}_+)\|_{L^p(X)} \lesssim_{d,k,p,\deg \mathcal{P}} \|f\|_{L^p(X)}. \tag{1.24}$$

Moreover, the implicit constants in (1.21), (1.22), (1.23) *and* (1.24) *can be taken to be independent of the coefficients of the polynomials from* $\mathcal{P}$, *depending only on* p *and the degree of the family* $\mathcal{P}$.

We now give some remarks about Theorem 1.20.

(1) Theorem 1.20 is a multi-dimensional, quantitative counterpart of Theorem 1.8 with sharp ranges of parameters $1 < p < \infty$ and $2 < r < \infty$, which contributes to the Furstenberg–Bergelson–Leibman conjecture (see Section 5.5, p. 468, in [4]) in the linear case for the class of commuting measure-preserving transformations. The Furstenberg–Bergelson–Leibman conjecture is a central open problem in pointwise ergodic theory. Moreover, inequalities (1.23) and (1.24) are the strongest possible quantitative forms of pointwise convergence. By taking $d = k = 1$ and $P_1(m) = m$ in Theorem 1.20, we recover Birkhoff's and von Neumann's results stated in Theorem 1.4. Taking $d = k = 1$ and $P_1 \in \mathbb{Z}[\mathrm{m}]$ in Theorem 1.20, we also recover Bourgain's polynomial ergodic theorem from Theorem 1.8 above.

(2) The mean ergodic theorem in (i) is a consequence of the dominated convergence theorem combined with (ii) and (iii). Each of the conclusions from (iv), (v) and (vi) individually implies pointwise convergence from (ii), as well as the maximal estimates from (iii). It also follows from (2.20) that (v) implies (iv). Details about these implications can be easily derived from the properties of oscillations, variations and jumps collected in Section 2.

(3) Sharp r-variational estimates (1.22) were obtained for the first time in [55], with a conceptually new proof which also works for other discrete operators with arithmetic features [56]. Not long afterwards, the ideas from [55] were extended [59] to establish uniform λ-jump estimates (1.23). Partial result for r-variational estimates (1.22) were obtained in [42, 60, 76].

(4) It was observed in [55] that (1.22) and Hölder's inequality imply that for every $p \in (1,\infty)$, for any $r > 2$, every $f \in L^p(X)$ and every $J \in \mathbb{Z}_+$, one has

$$\sup_{I\in\mathfrak{S}_J(\mathbb{Z}_+)} \|O^2_{I,J}(A^{\mathcal{P}}_{M;X,\mathcal{T}} f : M \in \mathbb{Z}_+)\|_{L^p(X)} \lesssim_{d,k,p,r,\deg \mathcal{P}} J^{1/2-1/r} \|f\|_{L^p(X)},$$

with the same implicit constant as in (1.22) and so blows up as r tends to 2. This inequality is a non-uniform version of (1.24) in the spirit of Bourgain's oscillation inequality (1.10). However it was observed recently [54] that the methods from [55, 59] give the uniform oscillation inequality in (1.24). From this point of view (and from the discussion above for martingales) inequality (1.24) can be thought of as an endpoint for (1.22) at $r = 2$, though it is not an endpoint in the sense of inequality (2.20) below. It would be nice to know whether it is possible (if at all) to use (1.24) to recover (1.22).

(5) Inequality (1.24) is also a contribution to an interesting problem from the early 1990's of Rosenblatt and Wierdl (Problem 4.12 on p. 80 of [70]) about uniform estimates of oscillation inequalities for ergodic averages. In [33], Jones, Kaufman, Rosenblatt and Wierdl proved (1.24) for the classical Birkhoff averages with $d = k = 1$ and $P_1(m) = m$, giving an affirmative answer to Problem 4.12 on p. 80 of [70]. In [54], it was shown that Problem 4.12 on p. 80 of [70] remains true even for multidimensional polynomial ergodic averages.

(6) The proof of Theorem 1.20 is an elaboration of methods developed in [55, 59] and also recently in [54]. The main tools are the Hardy–Littlewood circle method (major arcs estimates); Weyl's inequality (minor arcs estimates); the Ionescu–Wainger multiplier theory (see [32, 53, 59] and also [65], [74]); the Rademacher–Menshov argument (see for instance [58]); and the sampling principle of Magyar–Stein–Wainger (see [51] and also [57]). The methods from [53, 55, 57–59] were further developed by the first author in collaboration with Krause and Tao [43], which resulted in establishing pointwise convergence for the so-called bilinear Furstenberg–Weiss ergodic averages. This was a long-standing open problem, which makes a significant contribution towards the Furstenberg–Bergelson–Leibman conjecture [4].

1.7. A multi-parameter variant of the Bellow and Furstenberg problem

After completing [6–8], Bourgain observed that the Dunford–Zygmund theorem (see Theorem 1.5) can be extended to the polynomial setting at the expense of imposing that the measure-preserving transformations in Theorem 1.5 commute. Bourgain's result can be formulated as follows.

Theorem 1.25 (Polynomial Dunford–Zygmund ergodic theorem). *Let $d \in \mathbb{Z}_+$ and let $P_1, \dots, P_d \in \mathbb{Z}[\mathrm{m}]$ such that $P_j(0) = 0$ for $j \in [d]$ be given. Let $(X, \mathcal{B}(X), \mu)$ be a σ-finite measure space endowed with a family $\mathcal{T} = (T_1, \dots, T_d)$ of commuting invertible measure-preserving transformations on X. Let $f \in L^p(X)$ for some $1 \le p \le \infty$, and for $M \in \mathbb{Z}_+^d$, let $A_{M;X,\mathcal{T}}^{P_1(\mathrm{m}_1),\dots,P_d(\mathrm{m}_d)} f = A_{M_1,\dots,M_d;X,T_1,\dots,T_d}^{P_1(\mathrm{m}_1),\dots,P_d(\mathrm{m}_d)} f$ be the polynomial ergodic average defined in (1.1).*

- (i) (Mean ergodic theorem) *If $1 < p < \infty$, then the averages $A_{M;X,\mathcal{T}}^{P_1(\mathrm{m}_1),\dots,P_d(\mathrm{m}_d)} f$ converge in $L^p(X)$ norm as $\min\{M_1, \dots, M_d\} \to \infty$.*
- (ii) (Pointwise ergodic theorem) *If $1 < p < \infty$, then the averages $A_{M;X,\mathcal{T}}^{P_1(\mathrm{m}_1),\dots,P_d(\mathrm{m}_d)} f$ converge pointwise almost everywhere on X as $\min\{M_1, \dots, M_d\} \to \infty$.*

(iii) (Maximal ergodic theorem) *If* $1 < p \le \infty$*, then one has*

$$\Big\| \sup_{M\in\mathbb{Z}_+^d} |A_{M;X,\mathcal{T}}^{P_1(\mathrm{m}_1),\dots,P_d(\mathrm{m}_d)} f| \Big\|_{L^p(X)} \lesssim_{d,p,\deg P_1,\dots,\deg P_d} \|f\|_{L^p(X)}. \tag{1.26}$$

(iv) (Oscillation ergodic theorem) *If* $1 < p < \infty$*, then one has*

$$\sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{Z}_+^d)} \|O_{I,J}^2(A_{M;X,\mathcal{T}}^{P_1(\mathrm{m}_1),\dots,P_d(\mathrm{m}_d)} f : M \in \mathbb{Z}_+^d)\|_{L^p(X)} \lesssim_{d,p,\deg P_1,\dots,\deg P_d} \|f\|_{L^p(X)}. \tag{1.27}$$

(*We refer to Section* 2 *for the definitions of the sets* $\mathfrak{S}_J(\mathbb{Z}_+^d)$*, see* (2.2)*, and the multi-parameter oscillations, see* (2.3)*.) Moreover, the implicit constants in* (1.26) *and* (1.27) *can be taken to be independent of the coefficients of the polynomials* $P_1,\dots,P_d$*, depending only on* p *and* $\deg P_1,\dots,\deg P_d$.

We now give some remarks about Theorem 1.25.

(1) Theorem 1.25 (i)-(iii) is attributed to Bourgain, though it has never been published. The first and third authors learned about this result from Bourgain in October 2016, when they started to work with Bourgain and Stein on some aspects of multi-parameter ergodic theory [9].

(2) In this paper we prove Theorem 1.25 using a general abstract principle, see Proposition 4.1 in Section 4. In contrast to Bourgain's original observation, our proof of Theorem 1.25 relies on uniform bounds for multi-parameter oscillation inequalities.

(3) Theorem 1.25 (iv) with linear polynomials $P_1(m) = \cdots = P_d(m) = m$ was established in [35], where it was essential that $\mathcal{T} = (T_1,\dots,T_d)$ is a commuting family of measure-preserving transformations on X. It is straightforward to see that (iv) implies (iii) by (2.7), as well as (ii) by appealing to Proposition 2.8. Using the dominated convergence theorem with (ii) and (iii), we also obtain (i). So it suffices to prove (1.27), which we do in Section 4.

(4) To prove Theorem 1.25, it is essential to note that

$$A_{M;X,\mathcal{T}}^{P_1(\mathrm{m}_1),\dots,P_d(\mathrm{m}_d)} f = A_{M_1,\dots,M_d;X,T_1,\dots,T_d}^{P_1(\mathrm{m}_1),\dots,P_d(\mathrm{m}_d)} f = A_{M_1;X,T_1}^{P_1(\mathrm{m}_1)} \circ \dots \circ A_{M_d;X,T_d}^{P_d(\mathrm{m}_d)} f, \tag{1.28}$$

where the latter averages (defined in (1.7)) commute as long as the family $\mathcal{T} = (T_1,\dots,T_d)$ is commuting. Using identity (1.28) and iterating appropriately (1.24) with $k = d = 1$, we will be able to derive (1.27). We refer to Section 4 for details.

Theorem 1.25 can be thought of as a simple case of a multi-parameter variant of the Bellow and Furstenberg problem, which is a central open problem in modern ergodic theory, and can be subsumed under the following conjecture.

Conjecture 1.29. *Let* $d, k \in \mathbb{Z}_+$ *be given and let* $(X, \mathcal{B}(X), \mu)$ *be a probability measure space endowed with a family* $\mathcal{T} = (T_1,\dots,T_d)$ *of invertible commuting measure-preserving transformations on* X*. Assume that* $\mathcal{P} = (P_1,\dots,P_d) \subset \mathbb{Z}[\mathrm{m}_1,\dots,\mathrm{m}_k]$ *such that* $P_j(0) = 0$ *for* $j \in [d]$ *are given. Then for any* $f \in L^\infty(X)$*, the multi-parameter polynomial averages* $A_{M;X,\mathcal{T}}^{\mathcal{P}} f(x) = A_{M_1,\dots,M_k;X,T_1,\dots,T_d}^{P_1,\dots,P_d} f(x)$ *defined in* (1.1) *converge for* μ*-almost every* $x \in X$*, as* $\min\{M_1,\dots,M_k\} \to \infty$.

A few remarks about this conjecture, its history, and the current state of the art, are in order.

(1) As seen above, the case $d = k = 1$ of Conjecture 1.29 with $P_1(m) = m$ follows from Birkhoff's ergodic theorem, see Theorem 1.4. The case $d = k = 1$ of Conjecture 1.29 with arbitrary polynomials $P_1 \in \mathbb{Z}[\mathrm{n}]$ was the famous open problem of Bellow [1] and Furstenberg [23], and was solved by Bourgain [6–8] in the mid 1980's, see Theorem 1.8. The general case $d, k \in \mathbb{Z}_+$ of Conjecture 1.29 with arbitrary polynomials $P_1, \dots, P_d \in \mathbb{Z}[\mathrm{m}_1, \dots, \mathrm{m}_k]$ in the diagonal setting $M_1 = \cdots = M_k$, that is, the multi-dimensional one-parameter setting, follows from Theorem 1.20.

(2) A genuinely multi-parameter case $d = k \geq 2$ of Conjecture 1.29 for averages (1.1) with $P_j(m_1, \dots, m_d) = P_j(m_j)$, where $P_j \in \mathbb{Z}[\mathrm{m}_j]$ for $j \in [d]$ follows from Theorem 1.25, which extends the case of linear polynomials $P_1(m) = \cdots = P_d(m) = m$ established independently by Dunford [19] and Zygmund [77] in the early 1950's, see Theorem 1.5.

(3) Thanks to the product structure of (1.28), Theorem 1.5, as well as Theorem 1.25, have relatively simple one-parameter proofs, which are based on iterative applications of Theorem 1.4 and Theorem 1.20, respectively. This is explained in Proposition 4.1 below. However, the situation is dramatically different when orbits in (1.1) are defined along genuinely k-variate polynomials $P_1, \dots, P_d \in \mathbb{Z}[\mathrm{m}_1, \dots, \mathrm{m}_k]$ since then we lose the product structure (1.28). This can be illustrated by considering averages (1.1) for $d = 1$, $k = 2$ with, let us say, $P_1(m_1, m_2) = m_1^2 m_2^3$. Then Conjecture 1.29 becomes challenging. Surprisingly, even in this simple case, it seems that there is no simple way (like changing variables or interpreting the average from (1.1) as a composition of simpler one-parameter averages as in (1.28)) that would help us reduce the matter to the setup where pointwise convergence is known. This was one of the motivations leading to Conjecture 1.29.

(4) The Dunford–Zygmund theorem (see Theorem 1.5 above) was originally proved for not necessarily commuting, measure-preserving transformations $\mathcal{T} = (T_1, \dots, T_d)$ on X. However, it is well known for instance from the Bergelson–Leibman paper [4] that the commutation assumption imposed on the family $T_1, \dots, T_d : X \to X$ in (1.1) is essential in order to have an ergodic theorem if $\deg P_j \geq 2$ for at least one $j \in [d]$ and $d \geq 2$. Even in the one-parameter case (assuming $k = 1$) in (1.1), an ergodic theorem may fail. The question to what extent one can relax commutation relations among $T_1, \dots, T_d$ in (1.1), even in the one-parameter case, is very intriguing. This also motivates the desire to understand Conjecture 1.29 in the commutative setting first, as it is unclear whether Conjecture 1.29 is true for all polynomials $P_1, \dots, P_d \in \mathbb{Z}[\mathrm{m}_1, \dots, \mathrm{m}_k]$.

(5) With respect to the noncommutative setting, we mention that recently the first and second authors with Ionescu and Magyar [29] established Conjecture 1.29 with $k = 1$, $d \in \mathbb{Z}_+$ and arbitrary polynomials $P_1, \dots, P_d \in \mathbb{Z}[\mathrm{m}]$ in the diagonal nilpotent setting, i.e., one-parameter and multi-dimensional, when $\mathcal{T} = (T_1, \dots, T_d)$ is a family of invertible measure-preserving transformations of a σ-finite measure space $(X, \mathcal{B}(X), \mu)$ that generates a nilpotent group of step two. In view of the Bergelson–Leibman paper [4], the nilpotent setting is probably the most general setting where Conjecture 1.29 might be true, at least in the one-parameter case.

(6) We finally mention that progress towards establishing Conjecture 1.29 was recently made by the first and third authors in collaboration with Bourgain and Stein [9]. This

conjecture was verified for any integer $d \geq 2$ with $k = d - 1$ for averages (1.1) with polynomials

$$P_j(\mathrm{m}_1, \dots, \mathrm{m}_{d-1}) = \mathrm{m}_j \quad \text{for } j \in [d-1]; \quad \text{and} \\ P_d(\mathrm{m}_1, \dots, \mathrm{m}_{d-1}) = P(\mathrm{m}_1, \dots, \mathrm{m}_{d-1}), \tag{1.30}$$

whenever $P \in \mathbb{Z}[\mathrm{m}_1, \dots, \mathrm{m}_{d-1}]$ is a polynomial such that

$$P(0, \dots, 0) = \partial_1 P(0, \dots, 0) = \cdots = \partial_{d-1} P(0, \dots, 0) = 0,$$

which has partial degrees (as a polynomial of the variable m_i for any $i \in [d-1]$) at least two. Furthermore, it follows from [9] that for any $P \in \mathbb{Z}[\mathrm{m}_1, \dots, \mathrm{m}_d]$, the following averages,

$$A^P_{M;X,\mathcal{T}} f(x) := \mathbb{E}_{(m_1,\dots,m_d) \in Q_M} f(T^{P(m_1,\dots,m_d)} x), \quad x \in X, \tag{1.31}$$

where $M = (M_1, \dots, M_d) \in \mathbb{Z}^d_+$, do converge almost everywhere on X provided that $\min\{M_1, \dots, M_d\} \to \infty$. In fact, Conjecture 1.29 was originally formulated with averages (1.31); the authors learned about this from Jean Bourgain in a private communication in October 2016. The proof from [9] developed new methods from Fourier analysis and number theory. Even though the averages (1.1) with polynomials from (1.30) share a lot of difficulties that arise in the general case, there are some cases that are not covered by the methods developed in [9]. At this moment it is not clear whether Conjecture 1.29 is true in full generality. The work in [9] is a significant step towards understanding Conjecture 1.29 that sheds new light on the general case and will either lead to its full resolution or to a counterexample. The authors plan to investigate this question in the near future.

1.8. Overview of the paper

In this paper we prove an abstract principle for the so-called projective operators, see Theorem 3.1 in Section 3, which allows us to deal with one-parameter oscillation inequalities in a fairly unified way. As a consequence of Theorem 3.1, we give a simple proof of the Jones–Kaufman–Rosenblatt–Wierdl oscillation inequality for martingales (Theorem 6.4 on p. 930 of [33]), see Proposition 3.11, and then we prove oscillation inequalities for smooth bumps, see Proposition 3.13 and Theorem 3.17. Further, we discuss oscillation estimates for projection operators corresponding to orthonormal systems in Hilbert spaces, see Proposition 3.22, and finally we obtain new oscillations inequalities for the Carleson operator, see Proposition 3.34. In Section 4 we build a multi-parameter theory of oscillation estimates, see Proposition 4.1 and Corollary 4.6. As an application of our method, we give a simple proof of Theorem 1.25.

This paper can be viewed as a fairly systematic treatment of oscillation estimates in the one-parameter as well as multi-parameter settings in ergodic theory and analysis. In the multi-parameter setting, oscillation semi-norms seem to be the only viable tool that allows us to handle efficiently multi-parameter pointwise convergence problems. This is especially the case in [9] where operators with arithmetic features were studied. It also contrasts sharply with the one-parameter setting, where we have a variety of available tools to handle pointwise convergence problems: including oscillations, variations or jumps, see [37, 55] and the references given there.

2. Notation and useful tools

We now set some notation that will be used throughout the paper. Basic properties of one-parameter as well as multi-parameter r-oscillation semi-norms, r-variation semi-norms and λ-jump counting functions will be also gathered here. We borrow notation from [9], Section 2, and [54], Section 2.

2.1. Basic notation

Let $\mathbb{Z}_+ := \{1, 2, \ldots\}$, $\mathbb{N} := \{0, 1, 2, \ldots\}$ and $\mathbb{R}_+ := (0, \infty)$. For $d \in \mathbb{Z}_+$, the sets $\mathbb{Z}^d$, $\mathbb{R}^d$, $\mathbb{C}^d$ and $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$ have standard meaning. We will also consider the set of dyadic numbers $\mathbb{D} := \{2^n : n \in \mathbb{Z}\}$. For any $x \in \mathbb{R}$, we define the floor function

$$\lfloor x \rfloor := \max\{n \in \mathbb{Z} : n \leq x\}.$$

For $x, y \in \mathbb{R}$, let $x \wedge y := \min\{x, y\}$ and $x \vee y := \max\{x, y\}$. For every $N \in \mathbb{R}_+$ and $\mathbb{A} \subseteq \mathbb{R}$, define

$$[N] := (0, N] \cap \mathbb{Z} = \{1, \ldots, \lfloor N \rfloor\},$$

as well as

$$\mathbb{A}_{\leq N} := [0, N] \cap \mathbb{A}, \quad \mathbb{A}_{<N} := [0, N) \cap \mathbb{A}, \quad \mathbb{A}_{\geq N} := [N, \infty) \cap \mathbb{A}, \quad \mathbb{A}_{>N} := (N, \infty) \cap \mathbb{A}.$$

We use $\mathbb{1}_A$ to denote the indicator function of a set A. If S is a statement, we write $\mathbb{1}_S$ to denote its indicator, equal to 1 if S is true and 0 if S is false. For instance, $\mathbb{1}_A(x) = \mathbb{1}_{x \in A}$.

For two nonnegative quantities A and B, we write $A \lesssim B$ if there is an absolute constant $C > 0$ such that $A \leq CB$; however, $C > 0$ may change from occurrence to occurrence. We will write $A \simeq B$ when $A \lesssim B \lesssim A$. We will write $\lesssim_\delta$ or $\simeq_\delta$ to emphasize that the implicit constant depends on δ. For two functions $f : X \to \mathbb{C}$ and $g : X \to [0, \infty)$, we write $f = O(g)$ if there exists $C > 0$ such that $|f(x)| \leq Cg(x)$ for all $x \in X$. We will also write $f = O_\delta(g)$ if the implicit constant depends on δ.

2.2. Euclidean spaces

The standard inner product, the corresponding Euclidean norm, and the maximum norm on $\mathbb{R}^d$ are denoted, respectively, for any $x = (x_1, \ldots, x_d)$, $\xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d$, by

$$x \cdot \xi := \sum_{k=1}^{d} x_k \xi_k, \quad |x| := |x|_2 := \sqrt{x \cdot x}, \quad \text{and} \quad |x|_\infty := \max_{k \in [d]} |x_k|.$$

2.3. Function spaces

Throughout this paper, all vector spaces will be defined over $\mathbb{C}$. For a continuous linear map $T : B_1 \to B_2$ between two normed vector spaces B_1 and B_2, its operator norm will be denoted by $\|T\|_{B_1 \to B_2}$.

The triple $(X, \mathcal{B}(X), \mu)$ denotes a measure space X with a σ-algebra $\mathcal{B}(X)$ and a σ-finite measure μ. The space of all μ-measurable functions $f : X \to \mathbb{C}$ will be denoted by $L^0(X)$. The space of all functions in $L^0(X)$ whose modulus is integrable with p-th power is denoted by $L^p(X)$ for $p \in (0,\infty)$, whereas $L^\infty(X)$ denotes the space of all essentially bounded functions in $L^0(X)$. These notions can be extended to functions taking values in a separable normed vector space $(B, \|\cdot\|_B)$; for instance,

$$L^p(X;B) := \big\{F \in L^0(X;B) : \|F\|_{L^p(X;B)} := \|\|F\|_B\|_{L^p(X)} < \infty\big\},$$

where $L^0(X;B)$ denotes[1] the space of measurable functions from X to B (up to almost everywhere equivalence). For any $p \in [1,\infty]$, we define a weak-L^p space of measurable functions on X by setting

$$L^{p,\infty}(X) := \{f : X \to \mathbb{C} \colon \|f\|_{L^{p,\infty}(X)} < \infty\},$$

where for any $p \in [1,\infty)$ we have

$$\|f\|_{L^{p,\infty}(X)} := \sup_{\lambda>0} \lambda\mu(\{x \in X : |f(x)| > \lambda\})^{1/p}, \quad \text{and} \quad \|f\|_{L^{\infty,\infty}(X)} := \|f\|_{L^\infty(X)}.$$

In our case, we will mainly take $X = \mathbb{R}^d$ or $X = \mathbb{T}^d$ equipped with the Lebesgue measure, and $X = \mathbb{Z}^d$ endowed with the counting measure. If X is endowed with a counting measure, we will abbreviate $L^p(X)$ to $\ell^p(X)$, $L^p(X;B)$ to $\ell^p(X;B)$, and $L^{p,\infty}(X)$ to $\ell^{p,\infty}(X)$.

2.4. Fourier transform

We will use the convention that $\boldsymbol{e}(z) = e^{2\pi i z}$ for every $z \in \mathbb{C}$, where $i^2 = -1$. Let $\mathcal{F}_{\mathbb{R}^d}$ denote the Fourier transform on $\mathbb{R}^d$ defined for any $f \in L^1(\mathbb{R}^d)$ and for any $\xi \in \mathbb{R}^d$ as

$$\mathcal{F}_{\mathbb{R}^d} f(\xi) := \int_{\mathbb{R}^d} f(x)\, \boldsymbol{e}(x\cdot\xi)\, \mathrm{d}x.$$

We can also consider the Fourier transform for finite Borel measures σ on $\mathbb{R}^d$. If $f \in \ell^1(\mathbb{Z}^d)$, we define the discrete Fourier transform (Fourier series) $\mathcal{F}_{\mathbb{Z}^d}$, for any $\xi \in \mathbb{T}^d$, by setting

$$\mathcal{F}_{\mathbb{Z}^d} f(\xi) := \sum_{x\in\mathbb{Z}^d} f(x)\, \boldsymbol{e}(x\cdot\xi).$$

Sometimes we shall abbreviate $\mathcal{F}_{\mathbb{Z}^d} f$ or $\mathcal{F}_{\mathbb{R}^d} f$ to $\widehat{f}$, if the context will be clear.

Let $\mathbb{G} = \mathbb{R}^d$ or $\mathbb{G} = \mathbb{Z}^d$. It is well known that their corresponding dual groups are $\mathbb{G}^* = (\mathbb{R}^d)^* = \mathbb{R}^d$ or $\mathbb{G}^* = (\mathbb{Z}^d)^* = \mathbb{T}^d$, respectively. For any bounded function $\mathfrak{m}\colon \mathbb{G}^* \to \mathbb{C}$ and a test function $f\colon \mathbb{G} \to \mathbb{C}$, we define the Fourier multiplier operator by

$$T_{\mathbb{G}}[\mathfrak{m}]f(x) := \int_{\mathbb{G}^*} \boldsymbol{e}(-\xi\cdot x)\, \mathfrak{m}(\xi)\, \mathcal{F}_{\mathbb{G}} f(\xi)\, \mathrm{d}\xi, \quad \text{for } x \in \mathbb{G}. \tag{2.1}$$

One may think that $f\colon \mathbb{G} \to \mathbb{C}$ is a compactly supported function on $\mathbb{G}$ (and smooth if $\mathbb{G} = \mathbb{R}^d$) or any other function for which (2.1) makes sense.

[1] Note that there are various definitions of $L^0(X;B)$ in the literature.

2.5. Littlewood–Paley theory

Often we will control oscillation and variation semi-norms by certain square functions of the form

$$S(f)(x) := \Big(\sum_{k\in\mathbb{Z}} |\sigma_k * f(x)|^2\Big)^{1/2},$$

where $(\sigma_k)_{k\in\mathbb{Z}}$ is a sequence of Borel measures on $\mathbb{R}^d$ with bounded total variation satisfying $|\widehat{\sigma_k}(\xi)| \le C \min\{|a_{k+1}\xi|^{\alpha}, |a_k\xi|^{-\alpha}\}$ for some $\alpha > 0$ and for all $k \in \mathbb{Z}$. Here, $\inf_{k\in\mathbb{Z}} a_{k+1}/a_k > 1$. What we call *standard Littlewood–Paley arguments* sometimes refer to the arguments developed in the seminal paper [20]. In particular, Theorem B in [20] implies that the square function S satisfies L^p bounds $\|S(f)\|_{L^p} \le C_p\|f\|_{L^p}$ for all $p \in (1,\infty)$ whenever the corresponding maximal function σ^* associated to the measures $(\sigma_k)_{k\in\mathbb{Z}}$ satisfies the same L^p bounds.

At one point we will use a powerful square function bound of Rubio de Francia associated to any pairwise disjoint collection of intervals $(I_j : j \in \mathbb{Z})$ on $\mathbb{R}$. It states

$$\Big\|\Big(\sum_{j\in\mathbb{Z}} |T_{\mathbb{R}}[\mathbb{1}_{I_j}]f|^2\Big)^{1/2}\Big\|_{L^p(\mathbb{R})} \lesssim \|f\|_{L^p(\mathbb{R})}$$

whenever $p \in [2,\infty)$. See Theorem 1.2 in [71].

2.6. Coordinatewise order $\preceq$

For any $x = (x_1,\dots,x_k) \in \mathbb{R}^k$ and $y = (y_1,\dots,y_k) \in \mathbb{R}^k$, we say $x \preceq y$ if an only if $x_i \le y_i$ for each $i \in [k]$. We also write $x \prec y$ if and only if $x \preceq y$ and $x \ne y$, and $x \prec_{\mathrm{s}} y$ if and only if $x_i < y_i$ for each $i \in [k]$. Let $\mathbb{I} \subseteq \mathbb{R}^k$ be an index set such that $\#\mathbb{I} \ge 2$, and for every $J \in \mathbb{Z}_+ \cup \{\infty\}$ define the set

$$\mathfrak{S}_J(\mathbb{I}) := \big\{(t_i : i \in \mathbb{N}_{\le J}) \subseteq \mathbb{I}\colon t_0 \prec_{\mathrm{s}} t_1 \prec_{\mathrm{s}} \dots \prec_{\mathrm{s}} t_J\big\}, \tag{2.2}$$

where $\mathbb{N}_{\le\infty} := \mathbb{N}$. In other words, $\mathfrak{S}_J(\mathbb{I})$ is the family of all strictly increasing sequences (with respect to the coordinatewise order) of length $J+1$ taking their values in the set $\mathbb{I}$.

2.7. Oscillation semi-norms

Let $\mathbb{I} \subseteq \mathbb{R}^k$ be an index set such that $\#\mathbb{I} \ge 2$. Let $(\mathfrak{a}_t(x) : t \in \mathbb{I})$ be a k-parameter family of complex-valued measurable functions defined on X. For any $\mathbb{J} \subseteq \mathbb{I}$, any $1 \le r < \infty$ and a sequence $I = (I_i : i \in \mathbb{N}_{\le J}) \in \mathfrak{S}_J(\mathbb{I})$, the multi-parameter r-oscillation seminorm is defined by

$$O^r_{I,J}(\mathfrak{a}_t(x) : t \in \mathbb{J}) := \Big(\sum_{j=0}^{J-1} \sup_{t\in\mathbb{B}[I_j]\cap\mathbb{J}} |\mathfrak{a}_t(x) - \mathfrak{a}_{I_j}(x)|^r\Big)^{1/r}, \tag{2.3}$$

where $\mathbb{B}[I_i] := [I_{i1}, I_{(i+1)1}) \times \dots \times [I_{ik}, I_{(i+1)k})$ is a box determined by the element $I_i = (I_{i1},\dots,I_{ik})$ of the sequence $I \in \mathfrak{S}_J(\mathbb{I})$. In order to avoid problems with measurability, we always assume that $\mathbb{I} \ni t \mapsto \mathfrak{a}_t(x) \in \mathbb{C}$ is continuous for μ-almost every $x \in X$, or $\mathbb{J}$ is countable. We also use the convention that the supremum taken over the empty set is zero.

Remark 2.4. Let $1 \leq r < \infty$. Some remarks are in order.

(1) Clearly, $O^r_{I,J}(\mathfrak{a}_t : t \in \mathbb{J})$ defines a semi-norm.

(2) Let $\mathbb{I} \subseteq \mathbb{R}^k$ be an index set such that $\#\mathbb{I} \geq 2$, and let $\mathbb{J}_1, \mathbb{J}_2 \subseteq \mathbb{I}$ be disjoint. Then for any family $(\mathfrak{a}_t : t \in \mathbb{I}) \subseteq \mathbb{C}$, any $J \in \mathbb{Z}_+$ and any $I \in \mathfrak{S}_J(\mathbb{I})$, one has

$$O^r_{I,J}(\mathfrak{a}_t : t \in \mathbb{J}_1 \cup \mathbb{J}_2) \leq O^r_{I,J}(\mathfrak{a}_t : t \in \mathbb{J}_1) + O^r_{I,J}(\mathfrak{a}_t : t \in \mathbb{J}_2).$$

(3) Let $\mathbb{I} \subseteq \mathbb{R}^k$ be a countable index set such that $\#\mathbb{I} \geq 2$ and $\mathbb{J} \subseteq \mathbb{I}$. Then for any family $(\mathfrak{a}_t : t \in \mathbb{I}) \subseteq \mathbb{C}$, any $J \in \mathbb{Z}_+$, any $I \in \mathfrak{S}_J(\mathbb{I})$, one has

$$O^r_{I,J}(\mathfrak{a}_t : t \in \mathbb{J}) \lesssim \Big(\sum_{t \in \mathbb{I}} |\mathfrak{a}_t|^r\Big)^{1/r}. \tag{2.5}$$

(4) Let $(\mathfrak{a}_t : t \in \mathbb{I}^k)$ be a k-parameter family of measurable functions on X. For any $\mathbb{I} \subseteq \mathbb{R}$ with $\#\mathbb{I} \geq 2$ and any sequence $I = (I_i : i \in \mathbb{N}_{\leq J}) \in \mathfrak{S}_J(\mathbb{I})$ of length $J \in \mathbb{Z}_+ \cup \{\infty\}$, we define the diagonal sequence $\bar{I} = (\bar{I}_i : i \in \mathbb{N}_{\leq J}) \in \mathfrak{S}_J(\mathbb{I}^k)$ by setting $\bar{I}_i = (I_i, \dots, I_i) \in \mathbb{I}^k$ for each $i \in \mathbb{N}_{\leq J}$. Then for any $p \in [1, \infty]$ and for any $\mathbb{J} \subseteq \mathbb{I}^k$, one has

$$\sup_{I \in \mathfrak{S}_J(\mathbb{I})} \|O^r_{\bar{I},J}(\mathfrak{a}_t : t \in \mathbb{J})\|_{L^p(X)} \leq \sup_{I \in \mathfrak{S}_J(\mathbb{I}^k)} \|O^r_{I,J}(\mathfrak{a}_t : t \in \mathbb{J})\|_{L^p(X)}.$$

We now show that oscillation semi-norms always dominate maximal functions.

Proposition 2.6. *Assume that $k \in \mathbb{Z}_+$, $\mathbb{I} \subseteq \mathbb{R}$ is such that $\#\mathbb{I} \geq 2$, and let $(\mathfrak{a}_t : t \in \mathbb{I}^k)$ be a k-parameter family of measurable functions on X. Then for every $p \in [1, \infty]$ and $r \in [1, \infty)$, we have*

$$\Big\| \sup_{t \in (\mathbb{I} \setminus \{\sup \mathbb{I}\})^k} |\mathfrak{a}_t| \Big\|_{L^p(X)} \leq \sup_{t \in \mathbb{I}^k} \|\mathfrak{a}_t\|_{L^p(X)} + \sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{I})} \big\| O^r_{\bar{I},J}(\mathfrak{a}_t : t \in \mathbb{I}^k) \big\|_{L^p(X)}, \tag{2.7}$$

where $\bar{I} \in \mathfrak{S}_J(\mathbb{I}^k)$ is the diagonal sequence corresponding to a sequence $I \in \mathfrak{S}_J(\mathbb{I})$ as in Remark 2.4.

Proof. Let $a = \inf \mathbb{I}$ and $b = \sup \mathbb{I}$. We see that $a < b$, since $\#\mathbb{I} \geq 2$. We choose a decreasing sequence $(a_n : n \in \mathbb{N}) \subseteq \mathbb{I}$ and an increasing sequence $(b_n : n \in \mathbb{N}) \subseteq \mathbb{I}$ such that $a \leq a_n \leq b_n \leq b$ for every $n \in \mathbb{N}$, satisfying

$$\lim_{n \to \infty} a_n = a \quad \text{and} \quad \lim_{n \to \infty} b_n = b,$$

and such that $a_n = a$ for all $n \in \mathbb{N}$ if $a \in \mathbb{I}$. By the monotone convergence theorem, we get

$$\begin{aligned}
\Big\| \sup_{t \in (\mathbb{I} \setminus \{\sup \mathbb{I}\})^k} |\mathfrak{a}_t| \Big\|_{L^p(X)} &= \lim_{n \to \infty} \Big\| \sup_{t \in [a_n, b_n)^k \cap \mathbb{I}^k} |\mathfrak{a}_t| \Big\|_{L^p(X)} \\
&\leq \sup_n \|\mathfrak{a}_{\bar{a}_n}\|_{L^p(X)} + \sup_n \Big\| \sup_{t \in [a_n, b_n)^k \cap \mathbb{I}^k} |\mathfrak{a}_t - \mathfrak{a}_{\bar{a}_n}| \Big\|_{L^p(X)},
\end{aligned}$$

where $\bar{a}_n = (a_n, \dots, a_n) \in [a_n, b_n)^k \cap \mathbb{I}^k$, and consequently we obtain (2.7). ∎

A remarkable feature of the oscillation seminorms is that they imply pointwise convergence. This property is formulated precisely in the following proposition.

Proposition 2.8. *Let $(X, \mathcal{B}(X), \mu)$ be a σ-finite measure space. For $k \in \mathbb{Z}_+$, let $(\mathfrak{a}_t : t \in \mathbb{R}^k_+)$ be a k-parameter family of measurable functions on X. Suppose that there are $p, r \in [1, \infty)$ such that for any $J \in \mathbb{Z}_+$ one has*

$$\sup_{I \in \mathfrak{S}_J(\mathbb{R}_+)} \|O^r_{\bar{I},J}(\mathfrak{a}_t : t \in \mathbb{R}^k_+)\|_{L^p(X)} \le C_{p,r}(J),$$

where

$$\lim_{J\to\infty} J^{-1/(p\vee r)} C_{p,r}(J) = 0,$$

and $\bar{I} \in \mathfrak{S}_J(\mathbb{R}^k_+)$ is the diagonal sequence corresponding to a sequence $I \in \mathfrak{S}_J(\mathbb{R}_+)$ as in Remark 2.4. Then the limits

$$\lim_{\min\{t_1,\ldots,t_k\}\to\infty} \mathfrak{a}_{(t_1,\ldots,t_k)} \quad \textit{and} \quad \lim_{\max\{t_1,\ldots,t_k\}\to 0} \mathfrak{a}_{(t_1,\ldots,t_k)}, \tag{2.9}$$

exist μ-almost everywhere on X.

Proof. We only prove the first conclusion of (2.9), as the second one can be proved in much the same way. Suppose by contradiction that the first limit in (2.9) does not exist μ almost everywhere on X. Since μ is a σ-finite measure, then there exists $X_0 \subseteq X$ such that $\mu(X_0) < \infty$, and also there is a small $\delta > 0$ such that

$$\mu\big(\big\{x \in X_0 : \lim_{N\to\infty} \sup_{s,t\succeq\bar{N}} |\mathfrak{a}_s(x) - \mathfrak{a}_t(x)| > 2\delta\big\}\big) > 2\delta,$$

where $\bar{N} = (N, \ldots, N) \in \mathbb{Z}^k_+$. For $N \in \mathbb{Z}_+$, define

$$A_N := \big\{x \in X_0 : \sup_{s,t\succeq\bar{N}} |\mathfrak{a}_s(x) - \mathfrak{a}_t(x)| > 2\delta\big\}.$$

Note that $A_{N+1} \subseteq A_N$ for every $N \in \mathbb{Z}_+$, and consequently from the continuity of measure one has

$$\lim_{N\to\infty} \mu\big(\big\{x \in X_0 : \sup_{s,t\succeq\bar{N}} |\mathfrak{a}_s(x) - \mathfrak{a}_t(x)| > 2\delta\big\}\big) > 2\delta.$$

Hence there is an $N_0 \in \mathbb{Z}_+$ such that for every $N \ge N_0$, we have

$$\mu\big(\big\{x \in X_0 : \sup_{t\succeq\bar{N}} |\mathfrak{a}_t(x) - \mathfrak{a}_{\bar{N}}(x)| > \delta\big\}\big) > \delta.$$

For $M, N \in \mathbb{Z}_+$, we now define

$$B^N_M := \big\{x \in X_0 : \sup_{\bar{N}\preceq t\prec_s \bar{M}} |\mathfrak{a}_t(x) - \mathfrak{a}_{\bar{N}}(x)\big\} > \delta|.$$

We observe that $B^N_M \subseteq B^N_{M+1}$ for every $M, N \in \mathbb{Z}_+$ and using once again continuity of measure, we obtain for every $N \ge N_0$,

$$\lim_{M\to\infty} \mu(B^N_M) = \mu\big(\big\{x \in X_0 : \sup_{t\succeq\bar{N}} |\mathfrak{a}_t(x) - \mathfrak{a}_{\bar{N}}(x)| > \delta\big\}\big) > \delta. \tag{2.10}$$

Using (2.10) recursively, we can construct a strictly increasing sequence $(I_i : i \in \mathbb{N}) \subset \mathbb{R}_+$ with $I_0 = N_0$ such that for every $i \in \mathbb{N}$ we have

$$\mu\big(\big\{x \in X_0 : \sup_{t\in\mathbb{B}[\bar{I}_i]}|\mathfrak{a}_t(x) - \mathfrak{a}_{\bar{I}_i}(x)| > \delta\big\}\big) > \delta, \tag{2.11}$$

where $\bar{I}_i = (I_i, \ldots, I_i) \in \mathbb{R}^k_+$. Then by (2.11) we obtain for every $J \in \mathbb{Z}_+$ that

$$\begin{aligned} J\delta^{p+1} = \sum_{j=0}^{J-1} \delta^{p+1} &\le \int_X \sum_{j=0}^{J-1} \sup_{t\in\mathbb{B}[\bar{I}_j]} |\mathfrak{a}_t(x) - \mathfrak{a}_{\bar{I}_j}(x)|^p \, d\mu(x) \\ &\le J^{1-q/r} \sup_{I\in\mathfrak{S}_J(\mathbb{Z}_+)} \|O^r_{\bar{I},J}(\mathfrak{a}_t : t \in \mathbb{R}^k_+)\|^p_{L^p(X)}, \end{aligned}$$

where $q := p \wedge r$. Thus

$$J^{q/r}\delta^{p+1} \le \sup_{I\in\mathfrak{S}_J(\mathbb{R}_+)} \|O^r_{\bar{I},J}(\mathfrak{a}_t : t \in \mathbb{R}^k_+)\|^p_{L^p(X)} \le C_{p,r}(J)^p.$$

Letting $J \to \infty$ we get a contradiction. This completes the proof of Proposition 2.8. ∎

2.8. Variation semi-norms

We recall the definition of r-variations. For any $\mathbb{I} \subseteq \mathbb{R}$, any family $(\mathfrak{a}_t : t \in \mathbb{I}) \subseteq \mathbb{C}$, and any exponent $1 \le r < \infty$, the r-variation semi-norm is defined to be

$$V^r(\mathfrak{a}_t : t \in \mathbb{I}) := \sup_{J\in\mathbb{Z}_+} \sup_{\substack{t_0<\cdots<t_J \\ t_j\in\mathbb{I}}} \Big(\sum_{j=0}^{J-1} |\mathfrak{a}_{t_{j+1}} - \mathfrak{a}_{t_j}|^r\Big)^{1/r}, \tag{2.12}$$

where the latter supremum is taken over all finite increasing sequences in $\mathbb{I}$.

Remark 2.13. Some remarks about definition (2.12) are in order.

(1) Clearly, $V^r(\mathfrak{a}_t : t \in \mathbb{I})$ defines a semi-norm.

(2) The function $r \mapsto V^r(\mathfrak{a}_t : t \in \mathbb{I})$ is non-increasing. Moreover, if $\mathbb{I}_1 \subseteq \mathbb{I}_2$, then

$$V^r(\mathfrak{a}_t : t \in \mathbb{I}_1) \le V^r(\mathfrak{a}_t : t \in \mathbb{I}_2).$$

(3) Let $\mathbb{I} \subseteq \mathbb{R}$ be such that $\#\mathbb{I} \ge 2$. Let $(\mathfrak{a}_t : t \in \mathbb{R}) \subseteq \mathbb{C}$ be given, and let $r \in [1, \infty)$. If $V^r(\mathfrak{a}_t : t \in \mathbb{R}) < \infty$, then $\lim_{t\to\infty} \mathfrak{a}_t$ exists. Moreover, for any $t_0 \in \mathbb{I}$ one has

$$\sup_{t\in\mathbb{I}} |\mathfrak{a}_t| \le |\mathfrak{a}_{t_0}| + V^r(\mathfrak{a}_t : t \in \mathbb{I}). \tag{2.14}$$

(4) Let $\mathbb{I} \subseteq \mathbb{R}$ be such that $\#\mathbb{I} \ge 2$. Then for any $r \ge 1$, any family $(\mathfrak{a}_t : t \in \mathbb{I}) \subseteq \mathbb{C}$, any $J \in \mathbb{Z}_+ \cup \{\infty\}$, and any $I \in \mathfrak{S}_J(\mathbb{I})$, one has

$$O^r_{I,J}(\mathfrak{a}_t : t \in \mathbb{I}) \le V^r(\mathfrak{a}_t : t \in \mathbb{I}) \le 2\Big(\sum_{t\in\mathbb{I}} |\mathfrak{a}_t|^r\Big)^{1/r}. \tag{2.15}$$

(5) Let $(\mathfrak{a}_t(x) : t \in \mathbb{R}_+)$ be a family of complex-valued measurable functions on a σ-finite measure space $(X, \mathcal{B}(X), \mu)$. Then for any $p \geq 1$ and $r \geq 2$ we have

$$\sup_{N\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_N(\mathbb{R}_+)} \|O^r_{I,N}(\mathfrak{a}_t : t \in \mathbb{R}_+)\|_{L^p(X)} \tag{2.16}$$
$$\lesssim \sup_{N\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_N(\mathbb{D})} \|O^r_{I,N}(\mathfrak{a}_t : t \in \mathbb{D})\|_{L^p(X)} + \Big\|\Big(\sum_{n\in\mathbb{Z}} V^r(\mathfrak{a}_t : t \in [2^n, 2^{n+1}])^2\Big)^{1/2}\Big\|_{L^p(X)}.$$

The inequality (2.16) is an analogue of Lemma 1.3 on p. 6716 of [37] for oscillation semi-norms.

2.9. Jumps

The r-variation is closely related to the λ-jump counting function. Recall that for any $\lambda > 0$, the λ-jump counting function of a function $f : \mathbb{I} \to \mathbb{C}$ is defined by

$$\begin{aligned} N_\lambda f &:= N_\lambda(f(t) : t \in \mathbb{I}) \\ &:= \sup\big\{J \in \mathbb{N} : \exists_{\substack{t_0<\cdots<t_J \\ t_j\in\mathbb{I}}} : \min_{0\leq j\leq J-1} |f(t_{j+1}) - f(t_j)| \geq \lambda\big\}. \end{aligned} \tag{2.17}$$

Remark 2.18. Some remarks about definition (2.17) are in order.

(1) For any $\lambda > 0$ and a function $f : \mathbb{I} \to \mathbb{C}$, let us also define the following quantity:

$$\begin{aligned} \mathcal{N}_\lambda f &:= \mathcal{N}_\lambda(f(t) : t \in \mathbb{I}) \\ &:= \sup\big\{J \in \mathbb{N} : \exists_{\substack{s_1<t_1\leq\cdots\leq s_J<t_J \\ s_j,t_j\in\mathbb{I}}} : \min_{1\leq j\leq J} |f(t_j) - f(s_j)| \geq \lambda\big\}. \end{aligned}$$

Then one has $N_\lambda f \leq \mathcal{N}_\lambda f \leq N_{\lambda/2} f$.

(2) It is clear from these definitions that $f \mapsto \sup_{\lambda>0} \|\lambda N_\lambda(f(\cdot, t) : t \in \mathbb{I})^{1/\rho}\|_{L^p(X)}$ satisfies a quasi-triangle inequality. However it is not obvious whether a genuine triangle inequality is available for λ-jumps. In many applications, the problem can be overcome since there is always a comparable semi-norm in the following sense. Namely, for every $p \in (1,\infty)$, and $\rho \in (1,\infty)$ there exists a constant $0 < C < \infty$ such that for every measure space $(X, \mathcal{B}(X), \mu)$, and $\mathbb{I} \subseteq \mathbb{R}$, there exists a (subadditive) seminorm $|||\cdot|||$ such that the following two-sided inequality

$$C^{-1}|||f||| \leq \sup_{\lambda>0} \|\lambda N_\lambda(f(\cdot, t) : t \in \mathbb{I})^{1/\rho}\|_{L^p(X)} \leq C|||f|||$$

holds for all measurable functions $f : X \times \mathbb{I} \to \mathbb{C}$. This was established in Corollary 2.2 on p. 805 of [57].

(3) Let $(\mathfrak{a}_t(x) : t \in \mathbb{R})$ be a family of measurable functions on a σ-finite measure space $(X, \mathcal{B}(X), \mu)$. Let $\mathbb{I} \subseteq \mathbb{R}$ and $\#\mathbb{I} \geq 2$. Then for every $p \in [1,\infty]$ and $r \in [1,\infty)$, we have

$$\sup_{\lambda>0} \|\lambda N_\lambda(\mathfrak{a}_t : t \in \mathbb{I})^{1/r}\|_{L^p(X)} \leq \|V^r(\mathfrak{a}_t : t \in \mathbb{I})\|_{L^p(X)}, \tag{2.19}$$

since for all $\lambda > 0$ we have the following pointwise estimate:

$$\lambda N_\lambda(\mathfrak{a}_t(x) : t \in \mathbb{I})^{1/r} \leq V^r(\mathfrak{a}_t(x) : t \in \mathbb{I}).$$

(4) Let $(X, \mathcal{B}(X), \mu)$ be a σ-finite measure space and let $\mathbb{I} \subseteq \mathbb{R}$. Fix $p \in [1, \infty]$ and $1 \le \rho < r \le \infty$. Then for every measurable function $f : X \times \mathbb{I} \to \mathbb{C}$ we have the estimate

$$\|V^r\big(f(\cdot,t) : t \in \mathbb{I}\big)\|_{L^{p,\infty}(X)} \lesssim_{p,\rho,r} \sup_{\lambda>0} \|\lambda N_\lambda(f(\cdot,t) : t \in \mathbb{I})^{1/\rho}\|_{L^{p,\infty}(X)}. \tag{2.20}$$

The inequality (2.20) can be thought of as an inverse to inequality (2.19). A proof of (2.20) can be found in Lemma 2.3 on p. 805 of [57]. Moreover, one cannot replace $L^{p,\infty}(X)$ with $L^p(X)$ in (2.20), see Lemma 2.24 in [54]. One can also show that there is a function $f : \mathbb{Z}_+ \times \mathbb{Z}_+ \to \mathbb{R}$ such that

$$\sup_{N\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_N(\mathbb{Z}_+)} \|O^r_{I,N}(f(\cdot,n) : n \in \mathbb{Z}_+)\|_{\ell^p(\mathbb{Z}_+)} = \infty, \quad 2 \le r \le \infty,$$

but

$$\sup_{\lambda>0} \|\lambda N_\lambda(f(\cdot,n) : n \in \mathbb{Z}_+)^{1/2}\|_{\ell^p(\mathbb{Z}_+)} < \infty.$$

3. One-parameter oscillation estimates

We state a simple one-parameter oscillation estimate for projections, which has many interesting implications. Here we are inspired by observations of M. Lacey who highlighted and pointed out the importance of projections in pointwise ergodic theory; see [70].

Theorem 3.1. *Let $(X, \mathcal{B}(X), \mu)$ be a σ-finite measure space and let $\mathbb{I} \subseteq \mathbb{R}$ be such that $\#\mathbb{I} \ge 2$. Let $(P_t)_{t\in\mathbb{I}}$ be a family of projections; that is, the linear operators $P_t : L^0(X) \to L^0(X)$ satisfy*

$$P_s P_t = P_{s\wedge t}, \quad \textit{for } s \ne t. \tag{3.2}$$

If the set $\mathbb{I}$ is uncountable, then we assume in addition that $\mathbb{I} \ni t \mapsto P_t f$ is continuous μ-almost everywhere on X for every $f \in L^0(X)$. Let $p, r \in (1, \infty)$ be fixed. Suppose that the P_t are bounded on $L^p(X)$, and suppose that the following two estimates hold:

$$\sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{I})} \Big\|\Big(\sum_{j=0}^{J-1} |(P_{I_{j+1}} - P_{I_j})f|^r\Big)^{1/r}\Big\|_{L^p(X)} \lesssim_{p,r} \|f\|_{L^p(X)}, \quad f \in L^p(X), \tag{3.3}$$

and the vector-valued estimate uniformly in $(f_j)_{j\in\mathbb{Z}} \in L^p(X; \ell^r(\mathbb{Z}))$

$$\Big\|\Big(\sum_{j\in\mathbb{Z}} \sup_{t\in\mathbb{I}} |P_t f_j|^r\Big)^{1/r}\Big\|_{L^p(X)} \lesssim_{p,r} \Big\|\Big(\sum_{j\in\mathbb{Z}} |f_j|^r\Big)^{1/r}\Big\|_{L^p(X)}. \tag{3.4}$$

Then the following one-parameter oscillation estimate holds:

$$\sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{I})} \|O^r_{I,J}(P_t f : t \in \mathbb{I})\|_{L^p(X)} \lesssim_{p,r} \|f\|_{L^p(X)}, \quad f \in L^p(X). \tag{3.5}$$

Proof. Fix $J \in \mathbb{Z}_+$ and $I \in \mathfrak{S}_J(\mathbb{I})$ and observe, using (3.2), that

$$(P_t - P_{I_j})f = P_t(P_{I_{j+1}} - P_{I_j})f, \quad \text{whenever } I_j < t < I_{j+1}.$$

Using this identity and then (3.4), we see that

$$\begin{aligned}&\Big\|\Big(\sum_{j=0}^{J-1}\sup_{\substack{I_j<t<I_{j+1}\\ t\in\mathbb{I}}}|P_tf-P_{I_j}f|^r\Big)^{1/r}\Big\|_{L^p(X)}\\ &\le\Big\|\Big(\sum_{j=0}^{J-1}\sup_{t\in\mathbb{I}}|P_t(P_{I_{j+1}}-P_{I_j})f|^r\Big)^{1/r}\Big\|_{L^p(X)}\lesssim_{p,r}\Big\|\Big(\sum_{j=0}^{J-1}|(P_{I_{j+1}}-P_{I_j})f|^r\Big)^{1/r}\Big\|_{L^p(X)}.\end{aligned}$$

Now applying (3.3) we arrive at (3.5). The proof of Theorem 3.1 is complete. ■

Remark 3.6. A few remarks are in order.

(1) Theorem 3.1 will be applied mainly when $r=2$. Then the estimate in (3.3) is a square function estimate, which can be deduced from the estimate

$$\sup_{J\in\mathbb{Z}_+}\sup_{I\in\mathfrak{S}_J(\mathbb{I})}\sup_{\substack{|\varepsilon_j|\le 1\\ 0\le j\le J}}\Big\|\sum_{j=0}^{J-1}\varepsilon_j(P_{I_{j+1}}f-P_{I_j}f)\Big\|_{L^p(X)}\lesssim_p\|f\|_{L^p(X)},\quad f\in L^p(X). \tag{3.7}$$

In fact, the implication from (3.7) to (3.3) is a simple consequence of Khintchine's inequality.

(2) Let $(X,\mathcal{B}(X),\mu)$ be a σ-finite measure space, let $\mathbb{I}\subseteq\mathbb{R}$ be countable, and let $(T_t)_{t\in\mathbb{I}}$ be a family of bounded operators on $L^p(X)$ for $p\in(1,\infty)$ satisfying

$$\Big\|\Big(\sum_{t\in\mathbb{I}}|(T_t-P_t)f|^2\Big)^{1/2}\Big\|_{L^p(X)}\lesssim_p\|f\|_{L^p(X)},\quad f\in L^p(X), \tag{3.8}$$

where $(P_t)_{t\in\mathbb{I}}$ is a family of projections as in Theorem 3.1 satisfying (3.3) and (3.4) with $r=2$. Then one has

$$\sup_{J\in\mathbb{Z}_+}\sup_{I\in\mathfrak{S}_J(\mathbb{I})}\|O^2_{I,J}(T_tf:t\in\mathbb{I})\|_{L^p(X)}\lesssim_p\|f\|_{L^p(X)},\quad f\in L^p(X). \tag{3.9}$$

In fact, in view of (2.15), the inequality (3.8) easily reduces the 2-oscillation estimate for $(T_t)_{t\in\mathbb{I}}$ to a 2-oscillation estimate for $(P_t)_{t\in\mathbb{I}}$. This observation will be very useful in many applications. We will see how it works in the case of smooth bump functions, see Theorem 3.17.

(3) As we know, oscillation inequalities are important in pointwise convergence problems, and in the vast majority of applications it suffices to understand (3.9) for $p=2$. This can be nicely illustrated as follows: suppose for $p\in(1,\infty)$ one has an a priori maximal bound

$$\|\sup_{t\in\mathbb{I}}|P_tf|\|_{L^p(X)}\lesssim_p\|f\|_{L^p(X)},\quad f\in L^p(X). \tag{3.10}$$

Then (3.10) with $p=2$ can be used to verify (3.4) with $p=r=2$. Finally, it remains to verify (3.3) with $p=r=2$, which in many cases can be deduced by using Fourier techniques or exploiting almost-orthogonality phenomena invoking TT^* arguments, see Proposition 3.22.

We now derive some consequences of Theorem 3.1.

3.1. Oscillation inequalities for martingales

We recall some basic facts about martingales. We will follow notation from [28], Section 3, p. 165. Let $(X, \mathcal{B}(X), \mu)$ be a σ-finite measure space and let $\mathbb{I}$ be a totally ordered set. A sequence of sub-σ-algebras $(\mathcal{F}_t : t \in \mathbb{I})$ is called a *filtration* if it is increasing and the measure μ is σ-finite on each $\mathcal{F}_t$. A *martingale* adapted to a filtration $(\mathcal{F}_t : t \in \mathbb{I})$ is a family of functions $\mathfrak{f} = (\mathfrak{f}_t : t \in \mathbb{I}) \subseteq L^1(X, \mathcal{B}(X), \mu)$ such that $\mathfrak{f}_s = \mathbb{E}[\mathfrak{f}_t|\mathcal{F}_s]$ for every $s, t \in \mathbb{I}$ so that $s \le t$, where $\mathbb{E}[\cdot|\mathcal{F}]$ denotes the the conditional expectation operator with respect to a sub-σ-algebra $\mathcal{F} \subseteq \mathcal{B}(X)$. We say that a martingale $\mathfrak{f} = (\mathfrak{f}_t : t \in \mathbb{I}) \subseteq L^p(X, \mathcal{B}(X), \mu)$ is bounded if

$$\sup_{t\in\mathbb{I}} \|\mathfrak{f}_t\|_{L^p(X)} \lesssim_p 1.$$

Applying Theorem 3.1, we immediately recover the oscillation inequality of Jones–Kaufman–Rosenblatt–Wierdl [33], which in fact is an oscillation inequality for bounded martingales.

Proposition 3.11. *For every $p \in (1, \infty)$, there exists a constant $C_p > 0$ such that for every bounded martingale $\mathfrak{f} = (\mathfrak{f}_n : n \in \mathbb{Z}) \subseteq L^p(X, \mathcal{B}(X), \mu)$ corresponding to a filtration $(\mathcal{F}_n : n \in \mathbb{Z})$ one has*

$$\sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{Z})} \|O^2_{I,J}(\mathfrak{f}_n : n \in \mathbb{Z})\|_{L^p(X)} \le C_p \sup_{n\in\mathbb{Z}} \|\mathfrak{f}_n\|_{L^p(X)}. \tag{3.12}$$

Inequality (3.12) was established in Theorem 6.4 on p. 930 of [33]. The authors first established (3.12) for $p = 2$, then proved weak type $(1, 1)$ as well as $L^\infty \to \mathrm{BMO}$ variants of (3.12), and consequently derived (3.12) for all $p \in (1, \infty)$ by interpolation. Our approach is direct and will avoid using any interpolation arguments in the proof.

Proof of Proposition 3.11. Fix $p \in (1, \infty)$. Define projections by $P_n(f) := \mathbb{E}[f|\mathcal{F}_n]$ for any $n \in \mathbb{Z}$ and $f \in L^p(X)$. Since $\mathfrak{f} = (\mathfrak{f}_n : n \in \mathbb{Z})$ is a martingale, then $\mathfrak{f}_n = P_n(\mathfrak{f}_n)$ for any $n \in \mathbb{Z}$, and consequently (3.2) holds. Moreover, by Burkholder [11], see also [12], it is very well known that (3.7) holds, which in view of Remark 3.6 implies

$$\sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{Z}_+)} \Big\| \Big(\sum_{j=0}^{J-1} |P_{I_{j+1}}(\mathfrak{f}_{I_J}) - P_{I_j}(\mathfrak{f}_{I_J})|^2 \Big)^{1/2} \Big\|_{L^p(X)} \lesssim_p \sup_{n\in\mathbb{Z}} \|\mathfrak{f}_n\|_{L^p(X)}.$$

This consequently verifies inequality (3.3). Invoking the Fefferman–Stein inequality for non-negative submartingales (Theorem 3.2.7 on p. 178 of [28]), we obtain

$$\Big\| \Big(\sum_{j\in\mathbb{Z}} \sup_{n\in\mathbb{Z}} |\mathbb{E}[|f_j||\mathcal{F}_n]|^2 \Big)^{1/2} \Big\|_{L^p(X)} \lesssim_p \Big\| \Big(\sum_{j\in\mathbb{Z}} |f_j|^2 \Big)^{1/2} \Big\|_{L^p(X)},$$

uniformly in $(f_j)_{j\in\mathbb{Z}} \in L^p(X; \ell^2(\mathbb{Z}))$, which in turn verifies the vector-valued estimate from (3.4). Appealing to Theorem 3.1, the oscillation inequality (3.12) follows and the proof of Proposition 3.11 is complete. ∎

3.2. Oscillation inequalities for smooth bump functions

Our aim will be to show that oscillation inequalities hold for L^1-dilated smooth bump functions. We begin with the main estimate.

Proposition 3.13. *For $d \in \mathbb{Z}_+$, let $\chi\colon \mathbb{R}^d \to [0,1]$ be a smooth function satisfying*

$$\mathbb{1}_{[-1,1]^d} \le \chi \le \mathbb{1}_{[-2,2]^d} \quad \text{for } \xi \in \mathbb{R}^d. \tag{3.14}$$

For every $n \in \mathbb{Z}$ and $\xi \in \mathbb{R}^d$, define $\chi_{2^n}(\xi) := \chi(2^{-n}\xi)$. Then for every $p \in (1,\infty)$, one has

$$\sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{Z})} \|O^2_{I,J}(T_{\mathbb{R}^d}[\chi_{2^n}]f : n \in \mathbb{Z})\|_{L^p(\mathbb{R}^d)} \lesssim_p \|f\|_{L^p(\mathbb{R}^d)}, \tag{3.15}$$

uniformly in $f \in L^p(\mathbb{R}^d)$.

Proof. Setting $P_n f := T_{\mathbb{R}^d}[\chi_{2^n}]f$ for every $n \in \mathbb{Z}$, and using (3.14), one sees that P_n is a projection in the sense of (3.2). Standard arguments based on the Littlewood–Paley theory (see Section 2.5) show that (3.3) with $r = 2$ holds. By the Fefferman–Stein inequality [72], we also obtain (3.4). An application of Theorem 3.1 now gives (3.15) as desired. ∎

Now our aim will be to extend inequality (3.15) to continuous times and general smooth bump functions.

Remark 3.16. A few remarks concerning Proposition 3.13 are in order.

(1) An important feature of our approach in Proposition 3.13 is that we do not need to invoke the corresponding inequality for martingales in the proof. This stands in sharp contrast to variants of inequality (3.15) involving r-variations, where all arguments to the best of our knowledge use the corresponding r-variational inequalities for martingales.

(2) Of course, inequality (3.15) can be reduced to the martingale setting from Proposition 3.11 by invoking square function arguments (Lemma 3.2 on p. 6722 of [37]) and standard Littlewood–Paley theory. The details may be found in [58].

(3) With respect to the previous two remarks, it would be interesting to know whether the r-variational counterpart of Proposition 3.13 can be proved without appealing to r-variational inequalities for martingales, see Lépingle's inequality (1.13).

Theorem 3.17. *For $d \in \mathbb{Z}_+$, let $\phi\colon \mathbb{R}^d \to \mathbb{C}$ be a Schwartz function. For $t \in \mathbb{R}_+$ and $x \in \mathbb{R}^d$, define $\phi_t(x) := t^{-d}\phi(t^{-1}x)$. Then for every $p \in (1,\infty)$, one has*

$$\sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{R}_+)} \|O^2_{I,J}(\phi_t * f : t \in \mathbb{R}_+)\|_{L^p(\mathbb{R}^d)} \lesssim_p \|f\|_{L^p(\mathbb{R}^d)}, \quad f \in L^p(\mathbb{R}^d). \tag{3.18}$$

Remark 3.19. Theorem 3.17 immediately extends to families of partial convolution operators. If $\mathbb{R}^d = \mathbb{R}^n \times \mathbb{R}^m$, we write elements $x \in \mathbb{R}^d$ as $x = (x', x'')$, where $x' \in \mathbb{R}^n$ and $x'' \in \mathbb{R}^m$. Let ϕ be a Schwartz function on $\mathbb{R}^n$ and define

$$T_t f(x) = \int_{\mathbb{R}^n} f(x' - y, x'')\,\phi_t(y)\,dy.$$

The oscillation inequality (3.18) implies the corresponding oscillation inequality for the family of partial convolution operators $(T_t)_{t\in\mathbb{R}_+}$.

Proof of Theorem 3.17. To prove (3.18), in view of (2.16), it suffices to show

$$\sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{D})} \|O^2_{I,J}(\phi_t * f : t \in \mathbb{D})\|_{L^p(\mathbb{R}^d)} \lesssim_p \|f\|_{L^p(\mathbb{R}^d)}, \quad f \in L^p(\mathbb{R}^d), \tag{3.20}$$

and

$$\Big\| \Big(\sum_{k\in\mathbb{Z}} V^2\big(\phi_t * f : t \in [2^k, 2^{k+1}]\big)^2 \Big)^{1/2} \Big\|_{L^p(\mathbb{R}^d)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}, \quad f \in L^p(\mathbb{R}^d). \tag{3.21}$$

Short 2-variational estimates were treated in [37] and in particular, the estimate (3.21) follows directly from Lemma 6.1 in [37].

To establish (3.20), we first observe that we may assume that $\int_{\mathbb{R}^d} \phi(x)dx = 0$. Indeed, if $\int_{\mathbb{R}^d} \phi(x)dx \neq 0$, then by scaling we may assume that $\int_{\mathbb{R}^d} \phi(x)dx = \chi(0) = 1$, where χ appears in Proposition 3.13. By standard Littlewood–Paley arguments (see Section 2.5), we note that (3.8) holds with $T_t f = \phi_t * f$ and $P_t f = T_{\mathbb{R}^d}[\chi_t] f$. Thus, by Remark 3.6, we see that (3.20) follows from the oscillation inequality (3.15) and so we may assume ϕ has mean zero. Using (2.5), we see that

$$\text{LHS of (3.20)} \lesssim \Big\| \Big(\sum_{k\in\mathbb{Z}} |\phi_{2^k} * f|^2 \Big)^{1/2} \Big\|_{L^p(\mathbb{R}^d)} \lesssim \|f\|_{L^p(\mathbb{R}^d)};$$

the last inequality following directly from Theorem B in [20]; see Section 2.5. This completes the proof of Theorem 3.17. ∎

3.3. Oscillation inequalities for orthonormal systems

The following result justifies in a strong sense the importance of oscillation inequalities.

Proposition 3.22. *Let $(X, \mathcal{B}(X), \mu)$ be a σ-finite measure space such that the corresponding Hilbert space $L^2(X)$ is endowed with an orthonormal basis $(\Phi_n)_{n\in\mathbb{N}}$. Then the projection operators*

$$P_n f := \sum_{k=0}^{n} \langle f, \Phi_k \rangle \Phi_k, \quad f \in L^2(X), \tag{3.23}$$

satisfy the oscillation estimate

$$\sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{N}_{\le N})} \|O^2_{I,J}(P_n f : n \in \mathbb{N}_{\le N})\|_{L^2(X)} \lesssim \log(N+1)\, \|f\|_{L^2(X)}. \tag{3.24}$$

Furthermore, if the projection operators P_n satisfy the maximal estimate

$$\big\| \sup_{n\in\mathbb{N}} |P_n f| \big\|_{L^2(X)} \lesssim \|f\|_{L^2(X)}, \quad f \in L^2(X), \tag{3.25}$$

then one has the uniform bound

$$\sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{N})} \|O^2_{I,J}(P_n f : n \in \mathbb{N})\|_{L^2(X)} \lesssim \|f\|_{L^2(X)}, \quad f \in L^2(X). \tag{3.26}$$

Proof. It is easy to see that P_n from (3.23) satisfies (3.2). To verify (3.3), we fix $J \in \mathbb{Z}_+$ and $I \in \mathfrak{S}_J(\mathbb{N})$ and note that by orthogonality we have

$$\begin{aligned} \Big\| \Big(\sum_{j=0}^{J-1} |(P_{I_{j+1}} - P_{I_j}) f|^2 \Big)^{1/2} \Big\|_{L^2(X)}^2 &= \sum_{j=0}^{J-1} \sum_{k_1=I_j+1}^{I_{j+1}} \sum_{k_2=I_j+1}^{I_{j+1}} \langle f, \Phi_{k_1} \rangle \overline{\langle f, \Phi_{k_2} \rangle} \langle \Phi_{k_1}, \Phi_{k_2} \rangle \\ &\le \sum_{k\in\mathbb{N}} |\langle f, \Phi_k \rangle|^2 = \|f\|_{L^2(X)}^2, \end{aligned}$$

where in the last equality we have used Parseval's identity for orthonormal bases. This proves (3.3) with $p = r = 2$. A famous result of Rademacher [68] and Menshov [52] asserts that there is a constant $C > 0$ such that for any $N \in \mathbb{Z}_+$, the projection operator P_n from (3.23) satisfies

$$\Big\| \sup_{n\in[N]} |P_n f| \Big\|_{L^2(X)} \le C \log(N+1) \Big(\sum_{n\in[N]} |\langle f, \Phi_n\rangle|^2 \Big)^{1/2} \lesssim \log(N+1) \|f\|_{L^2(X)}. \tag{3.27}$$

Using (3.27), we see that (3.4) holds with $p = r = 2$ with constant $\log(N + 1)$. Now applying Theorem 3.1 we obtain (3.24).

Under condition (3.25), we see that (3.4) holds with a uniform constant for $p = r = 2$ and so, applying Theorem 3.1 again, we obtain (3.26). ■

Proposition 3.22 is a key example in the study of oscillation semi-norms from the point of view their importance and usefulness in pointwise convergence problems. It exhibits, in view of inequality (2.7), that oscillation estimates (3.26) and maximal estimates (3.25) are equivalent in the class of orthonormal systems.

However, we have to emphasize that the maximal estimate from (3.25) is a very strong condition. On the one hand, we have Menshov's construction [52] of an orthonormal basis $(\Psi_n)_{n\in\mathbb{N}} \subseteq L^2([0,1])$ and a function $f_0 \in L^2([0,1])$ with almost everywhere diverging partial sums $\sum_{k=0}^n \langle f, \Psi_k\rangle \Psi_k$. Therefore maximal estimate (3.25) for Menshov's system cannot hold. In fact, the best what we can expect in the general case is the Rademacher–Menshov bound (3.27). The above-mentioned Menshov's construction [52] also shows that (3.27) is sharp and that the logarithm in (3.27) cannot be removed.

On the other hand, there is the famous result of Carleson (see [15]) which led to establishing (3.25) for the canonical trigonometric system $(\boldsymbol{e}(n\xi))_{n\in\mathbb{Z}}$ on $L^2([0,1])$ (see also [22, 27, 48]).

3.4. Oscillation inequalities for the Carleson operator

In this subsection, we obtain certain r-oscillation estimates for partial Fourier integrals on the real line $\mathbb{R}$.

The Carleson operator $\mathcal{C}_t$ is defined, for $f \in \mathcal{S}(\mathbb{R})$, $x \in \mathbb{R}$ and $t \in \mathbb{R}_+$, by

$$\mathcal{C}_t f(x) := T_{\mathbb{R}}[\mathbb{1}_{[-t,t]}] f(x) = \int_{-t}^{t} \mathcal{F}_{\mathbb{R}} f(\xi)\, \boldsymbol{e}(-x\xi)\, d\xi, \tag{3.28}$$

The celebrated Carleson–Hunt theorem (see the papers of Carleson [15] and Hunt [27]) asserts that for every $p \in (1,\infty)$ there is a constant $C_p > 0$ such that

$$\Big\| \sup_{t>0} |\mathcal{C}_t f| \Big\|_{L^p(\mathbb{R})} \le C_p \|f\|_{L^p(\mathbb{R})}, \quad f \in L^p(\mathbb{R}). \tag{3.29}$$

Remark 3.30. A few remarks about the Carleson–Hunt theorem are in order.

(1) Carleson [15] originally proved that the maximal partial sum operator of Fourier series corresponding to square-integrable functions on the circle is weak type $(2, 2)$. Not long afterwards this result was extended by Hunt [27] who proved that the maximal partial sum operator of Fourier series is bounded on $L^p(\mathbb{T})$ for any $p \in (1,\infty)$.

(2) Kenig and Tomas [39] used a transplantation arugment to show that the latter result is equivalent to inequality (3.29). This equivalence was extended to variation and oscillation inequalities in [63]. The foundational work of Kolmogorov [40, 41] shows that the range of $p \in (1, \infty)$ in inequality (3.29) is sharp.

(3) An alternative proof of Carleson's theorem was provided by Fefferman [22], who pioneered the ideas of the so called time-frequency analysis.

(4) Lacey and Thiele [48] established an independent proof on the real line of the weak type $(2, 2)$ boundedness of the maximal Fourier integral operator (3.28). The latter bound was extended by Grafakos, Tao, and Terwilleger [25] to (3.29) for all $p \in (1, \infty)$, see also [67].

(5) Inequality (3.29) was extended to the vector-valued setting by Grafakos, Martell and Soria [24], who proved that that for every $p, r \in (1, \infty)$, there is a constant $C_{p,r} > 0$ such that

$$\Big\|\Big(\sum_{j\in\mathbb{Z}}\sup_{t>0}|\mathcal{C}_t f_j|^r\Big)^{1/r}\Big\|_{L^p(\mathbb{R})} \le C_{p,r}\Big\|\Big(\sum_{j\in\mathbb{Z}}|f_j|^r\Big)^{1/r}\Big\|_{L^p(\mathbb{R})}, \tag{3.31}$$

uniformly in $(f_j)_{j\in\mathbb{Z}} \in L^p(X; \ell^r(\mathbb{Z}))$.

(6) We finally refer to the survey of Lacey [45], where details (including comprehensive historical background) and an extensive literature are given about this fascinating subject of pointwise convergence of Fourier series and related topics.

A far-reaching quantitative extension of (3.29) was obtained by the third author in collaboration with Oberlin, Seeger, Tao and Thiele [63], which asserts that for every $p \in (1, \infty)$ and for every $r > \max\{2, p/(p-1)\}$, there is a constant $C_{p,r} > 0$ such that

$$\|V^r(\mathcal{C}_t f : t \in \mathbb{R}_+)\|_{L^p(\mathbb{R})} \le C_{p,r}\|f\|_{L^p(\mathbb{R})}, \quad f \in L^p(\mathbb{R}). \tag{3.32}$$

See also in [75] for a different proof using outer measures. Furthermore, a restricted weak-type bound is established at the endpoint $p = r'$ when $p \in (1, 2)$ (here $r' = r/(r-1)$) and it is open whether weak type (p, p) holds true. It also follows from [63] that the ranges of parameter $p \in (1, \infty)$ and $r > \max\{2, p'\}$ in (3.32) are sharp. In the endpoint case $p = r'$, the Lorentz space $L^{r',\infty}$ cannot be replaced by a smaller Lorentz space. For weighted variational estimates for the Carleson operator, see [18] and [17], and the references given there.

Inequality (3.32), in view of inequality (2.15), immediately implies that for every $p \in (1, \infty)$ and for every $r > \max\{2, p'\}$, there is a constant $C_{p,r} > 0$ (actually, the same as in (3.32)) such that

$$\sup_{J\in\mathbb{Z}_+}\sup_{I\in\mathfrak{S}_J(\mathbb{R}_+)}\|O^r_{I,J}(\mathcal{C}_t f : t \in \mathbb{R}_+)\|_{L^p(\mathbb{R})} \le C_{p,r}\|f\|_{L^p(\mathbb{R})}, \quad f \in L^p(\mathbb{R}). \tag{3.33}$$

For applications of (3.32) and (3.33) to the Wiener–Wintner theorem in ergodic theory, see [47] and [63].

It has been observed by M. Lacey [44] (see [70] for the case $p = 2$) that (3.33) remains true for $r = 2$ whenever $p \in [2, \infty)$. Furthermore, this can be extend to all $p > 1$ when we restrict the t parameter in $\mathcal{C}_t$ to dyadic numbers $t \in \mathbb{D}$. Our aim here is to show how these results follow as an immediate consequence of Theorem 3.1.

Proposition 3.34. *Let $(\mathcal{C}_t)_{t\in\mathbb{R}_+}$ be as in (3.28). Then for every $p \in [2,\infty)$, there exists a constant $C_p > 0$ such that*

$$\sup_{J\in\mathbb{Z}_+}\sup_{I\in\mathfrak{S}_J(\mathbb{R}_+)}\|O^2_{I,J}(\mathcal{C}_t f : t\in\mathbb{R}_+)\|_{L^p(\mathbb{R})} \le C_p\|f\|_{L^p(\mathbb{R})}, \quad f\in L^p(\mathbb{R}). \tag{3.35}$$

Furthermore, for $(\mathcal{C}_t)_{t\in\mathbb{D}}$, we have

$$\sup_{J\in\mathbb{Z}_+}\sup_{I\in\mathfrak{S}_J(\mathbb{D})}\|O^2_{I,J}(\mathcal{C}_t f : t\in\mathbb{D})\|_{L^p(\mathbb{R})} \le C_p\|f\|_{L^p(\mathbb{R})}, \quad \textit{for all } p\in(1,\infty). \tag{3.36}$$

Proof. Observe the operators $(\mathcal{C}_t)_{t\in\mathbb{R}_+}$ are projections in the sense of (3.2). Moreover, when the sequence $(I_j)_{j\in\mathbb{N}}\subset\mathbb{D}$ lies among the dyadic numbers, the bound

$$\sup_{J\in\mathbb{Z}_+}\sup_{I\in\mathfrak{S}_J(\mathbb{D})}\Big\|\Big(\sum_{j=0}^{J-1}|(\mathcal{C}_{I_{j+1}}-\mathcal{C}_{I_j})f|^2\Big)^{1/2}\Big\|_{L^p(\mathbb{R})} \lesssim_p \|f\|_{L^p(\mathbb{R})}, \quad p\in(1,\infty),$$

follows from the classical Littlewood–Paley inequality associated to dyadic intervals (no need to refer to the refinements of the theory from Section 2.5). This verifies (3.3) with $r=2$ and $p\in(1,\infty)$ in the dyadic case. Furthermore, by Rubio de Francia's square function theorem for intervals (see Section 2.5), one has for every $p\in[2,\infty)$ that

$$\sup_{J\in\mathbb{Z}_+}\sup_{I\in\mathfrak{S}_J(\mathbb{R}_+)}\Big\|\Big(\sum_{j=0}^{J-1}|(\mathcal{C}_{I_{j+1}}-\mathcal{C}_{I_j})f|^2\Big)^{1/2}\Big\|_{L^p(\mathbb{R})} \lesssim_p \|f\|_{L^p(\mathbb{R})}, \quad f\in L^p(\mathbb{R}),$$

which verifies (3.3) with $r=2$ and $p\in[2,\infty)$. Using (3.31) with $r=2$, we also see that (3.4) is verified with $r=2$ and $p\in(1,\infty)$. Thus, invoking Theorem 3.1, inequalities (3.35) and (3.36) follow. ■

Proposition 3.34 for $p=2$ was established by Rosenblatt and Wierdl (see inequality (4.12) on p. 82 of [70]). In [47], Lacey and Terwilleger established (3.36) for $p\in(1,\infty)$. Proposition 3.34 gives a simple proof of these results.

In view of inequality (2.7) it is not difficult to see that the maximal estimates (3.29) and the oscillation estimates (3.35) for the Carleson operator are equivalent for all $p\in[2,\infty)$.

We also remark that the proof above also gives a proof of (3.33) which does not appeal to the variational inequality (3.32). Indeed, Rubio de Francia's result (inequality (7.1) on p. 10 of [71]) states that for every $p\in(1,2)$ and $r>p'$, one has

$$\sup_{J\in\mathbb{Z}_+}\sup_{I\in\mathfrak{S}_J(\mathbb{R}_+)}\Big\|\Big(\sum_{j=0}^{J-1}|(\mathcal{C}_{I_{j+1}}-\mathcal{C}_{I_j})f|^r\Big)^{1/r}\Big\|_{L^p(\mathbb{R})} \lesssim_{p,r} \|f\|_{L^p(\mathbb{R})}, \tag{3.37}$$

for $f\in L^p(\mathbb{R})$. Hence using (3.37) and (3.31) and invoking Theorem 3.1, we obtain the desired claim in (3.33).

A counterexample of Cowling and Tao [16] to Rubio de Francia's conjecture (Conjecture 7.2 in [71]) shows that for all $p\in(1,2)$, one has

$$\sup_{\|f\|_{L^p(\mathbb{R})}\le 1}\sup_{I\in\mathfrak{S}_\infty(\mathbb{R}_+)}\Big\|\Big(\sum_{j=0}^{\infty}|(\mathcal{C}_{I_{j+1}}-\mathcal{C}_{I_j})f|^r\Big)^{1/r}\Big\|_{L^p(\mathbb{R})} = \infty, \quad \text{where} \quad r=\frac{p}{p-1}.$$

Therefore (3.33) for $r=p'$ with $p\in(1,2)$ cannot hold. This shows that the range of p and r in (3.33) and (3.35) is sharp.

4. Multi-parameter oscillation estimates

In this section we establish Theorem 1.25. We begin with proving an abstract multi-parameter oscillation result, which may be of independent interest. Before we do this, we need more notation. For a linear operator $T\colon L^0(X) \to L^0(X)$, we shall denote by $|T|$ the sublinear maximal operator taken in the lattice sense defined by

$$|T|f(x) = \sup_{|g|\le|f|} |Tg(x)|, \quad x \in X, \text{ and } f \in L^p(X).$$

For two linear operators $S, T\colon L^0(X) \to L^0(X)$, we have $|ST|f \le |S||T|f$ whenever $f \in L^0(X)$.

Proposition 4.1. *Let $(X, \mathcal{B}(X), \mu)$ be a σ-finite measure space and let $\mathbb{I} \subseteq \mathbb{R}$ be such that $\#\mathbb{I} \ge 2$. Let $k \in \mathbb{N}_{\ge 2}$ and $p, r \in (1,\infty)$ be fixed. Let $(T_t)_{t\in\mathbb{I}^k}$ be a family of linear operators of the form*

$$T_t := T^1_{t_1} \cdots T^k_{t_k}, \quad t = (t_1, \dots, t_k) \in \mathbb{I}^k,$$

where $\{T^i_{t_i} : i \in [k],\ t_i \in \mathbb{I}\}$ is a family of commuting linear operators, which are bounded on $L^p(X)$. If the set $\mathbb{I}$ is uncountable, then we also assume that $\mathbb{I} \ni t \mapsto T^i_t f$ is continuous μ-almost everywhere on X for every $f \in L^0(X)$ and $i \in [k]$. Further assume that for every $i \in [k]$, we have

$$\sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{I})} \|O^r_{I,J}(T^i_t f : t \in \mathbb{I})\|_{L^p(X)} \lesssim_{p,r} \|f\|_{L^p(X)}, \quad f \in L^p(X), \tag{4.2}$$

and

$$\Big\|\Big(\sum_{j\in\mathbb{Z}}\big(\sup_{t\in\mathbb{I}}|T^i_t||f_j|\big)^r\Big)^{1/r}\Big\|_{L^p(X)} \lesssim_{p,r} \Big\|\Big(\sum_{j\in\mathbb{Z}}|f_j|^r\Big)^{1/r}\Big\|_{L^p(X)}, \tag{4.3}$$

uniformly in $(f_j)_{j\in\mathbb{Z}} \in L^p(X;\ell^r(\mathbb{Z}))$. Then we have the multi-parameter r-oscillation estimate

$$\sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{I}^k)} \|O^r_{I,J}(T_t f : t \in \mathbb{I}^k)\|_{L^p(X)} \lesssim \|f\|_{L^p(X)}, \quad f \in L^p(X).$$

Proof. For $i \in [k]$ and $n = (n_1, \dots, n_{i-1}, n_{i+1}, \dots, n_k) \in \mathbb{I}^{k-1}$, let us denote

$$T^{(i)}_n := T^1_{n_1} \cdots T^{i-1}_{n_{i-1}} T^{i+1}_{n_{i+1}} \cdots T^k_{n_k}.$$

Using this definition, the bound (4.3) and proceeding inductively, we easily see that

$$\Big\|\Big(\sum_{j\in\mathbb{Z}}\big(\sup_{n\in\mathbb{I}^{k-1}}|T^{(i)}_n||f_j|\big)^r\Big)^{1/r}\Big\|_{L^p(X)} \lesssim_{p} \Big\|\Big(\sum_{j\in\mathbb{Z}}|f_j|^r\Big)^{1/r}\Big\|_{L^p(X)}, \tag{4.4}$$

uniformly in $i \in [k]$ and $(f_j)_{j\in\mathbb{Z}} \in L^p(X;\ell^r(\mathbb{Z}))$. Furthermore, for $n \in \mathbb{I}^k$ and $I_j = (I_{j1}, \dots, I_{jk}) \in \mathbb{I}^k$, we have the identity

$$T_n f - T_{I_j} f = \sum_{m=1}^{k} T^{(m)}_{n(m,n,I_j)} (T^m_{n_m} - T^m_{I_{jm}}) f, \tag{4.5}$$

where $n(m, n, I_j) := (n_1, \dots, n_{m-1}, I_{j(m+1)}, \dots, I_{jk}) \in \mathbb{I}^{k-1}$.

We now fix $J \in \mathbb{Z}_+$ and a sequence $I \in \mathfrak{S}_J(\mathbb{I}^k)$. Applying the identity (4.5), the triangle inequality, the bound (4.4) applied to

$$f_j^m = \sup_{\substack{I_{jm} \le n_m < I_{(j+1)m} \\ n_m \in \mathbb{I}}} |T_{n_m}^m f - T_{I_{jm}}^m f|$$

and (4.2), we obtain

$$\begin{aligned}
&\Big\| \Big(\sum_{j=0}^{J-1} \sup_{n \in \mathbb{B}[I_j] \cap \mathbb{I}^k} |T_n f - T_{I_j} f|^r \Big)^{1/r} \Big\|_{L^p(X)} \\
&\quad \le \sum_{m=1}^{k} \Big\| \Big(\sum_{j=0}^{J-1} \Big(\sup_{n \in \mathbb{B}[I_j] \cap \mathbb{I}^k} |T_{n(m,n,I_j)}^{(m)}| \, |T_{n_m}^m f - T_{I_{jm}}^m f| \Big)^r \Big)^{1/r} \Big\|_{L^p(X)} \\
&\quad \le \sum_{m=1}^{k} \Big\| \Big(\sum_{j=0}^{J-1} \Big(\sup_{n \in \mathbb{I}^{k-1}} |T_n^{(m)}| \Big(\sup_{\substack{I_{jm} \le n_m < I_{(j+1)m} \\ n_m \in \mathbb{I}}} |T_{n_m}^m f - T_{I_{jm}}^m f| \Big) \Big)^r \Big)^{1/r} \Big\|_{L^p(X)} \\
&\quad \lesssim \sum_{m=1}^{k} \Big\| \Big(\sum_{j=0}^{J-1} \sup_{\substack{I_{jm} \le n_m < I_{(j+1)m} \\ n_m \in \mathbb{I}}} |T_{n_m}^m f - T_{I_{jm}}^m f|^r \Big)^{1/r} \Big\|_{L^p(X)} \lesssim \|f\|_{L^p(X)}.
\end{aligned}$$

This completes the proof of Proposition 4.1. ∎

We have a simple consequence of the above result.

Corollary 4.6. *Let $k \in \mathbb{N}_{\ge 2}$ and fix parameters $n_1, \dots, n_k \in \mathbb{Z}_+$, and $p \in (1, \infty)$. For every $i \in [k]$, let $\phi^i : \mathbb{R}^{n_i} \to \mathbb{C}$ be a Schwartz function, and define $\phi_{t_i}^i(x) := t_i^{-n_i} \phi^i(t_i^{-1} x)$ for every $t_i \in \mathbb{R}_+$ and $x \in \mathbb{R}^{n_i}$. Set $N := n_1 + \dots + n_k$ and for $t = (t_1, \dots, t_k) \in \mathbb{R}_+^k$ and $x = (x_1, \dots, x_k) \in \mathbb{R}^N := \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_k}$, consider the operator $T_t \colon L^p(\mathbb{R}^N) \to L^p(\mathbb{R}^N)$ defined by*

$$T_t f(x) := \int_{\mathbb{R}^{n_1}} \cdots \int_{\mathbb{R}^{n_k}} \Big(\prod_{i=1}^{k} \phi_{t_i}^i(z_i) \Big) f(x - z) \, dz_1 \dots dz_k, \quad z = (z_1, \dots, z_k).$$

Then we have the following multi-parameter oscillation estimate:

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{R}_+^k)} \|O_{I,J}^2(T_t f : t \in \mathbb{R}_+^k)\|_{L^p(\mathbb{R}^N)} \lesssim_p \|f\|_{L^p(\mathbb{R}^N)}, \quad f \in L^p(\mathbb{R}^N). \tag{4.7}$$

Proof. For $i \in [k]$ and $z_i \in \mathbb{R}^{n_i}$, we denote by $z_i^{(i)} = (z_1^{(i)}, \dots, z_k^{(i)})$ the point in $\mathbb{R}^N = \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_k}$ such that $z_j^{(i)} = \mathbb{1}_{\{j\}}(i) z_i \in \mathbb{R}^{n_i}$ for any $j \in [k]$. We define the operators $T_{t_i}^i \colon L^p(\mathbb{R}^N) \to L^p(\mathbb{R}^N)$ by

$$T_{t_i}^i f(x) := \int_{\mathbb{R}^{n_i}} \phi_{t_i}^i(z_i) f(x - z_i^{(i)}) \, dz_i, \quad x = (x_1, \dots, x_k) \in \mathbb{R}^N, t_i \in \mathbb{R}_+.$$

These operators commute and we have $T_t = T^1_{t_1} \circ \cdots \circ T^k_{t_k}$. Furthermore, these are partial convolution operators with Schwartz functions and so Theorem 3.17 (see Remark 3.19) implies that the oscillation estimate (4.2) holds for the family $(T^i_t)_{t\in\mathbb{R}_+}$, for each $i \in [k]$. Finally, the Fefferman–Stein vector-valued maximal inequality shows that (4.3) holds and so Proposition 4.1 gives us the desired conclusion (4.7). This completes the proof of Corollary 4.6. ∎

We close this section by establishing the main ergodic result of this survey.

Proof of Theorem 1.25. We will invoke Proposition 4.1 with $k = d$ and $r = 2$. As in (1.28) note that

$$A^{P_1(\mathrm{m}_1),\dots,P_d(\mathrm{m}_d)}_{M;X,\mathcal{T}} f = A^{P_1(\mathrm{m}_1),\dots,P_d(\mathrm{m}_d)}_{M_1,\dots,M_d;X,T_1,\dots,T_d} f = A^{P_1(\mathrm{m}_1)}_{M_1;X,T_1} \circ \cdots \circ A^{P_d(\mathrm{m}_d)}_{M_d;X,T_d} f,$$

where the averages $A^{P_1(\mathrm{m}_1)}_{M_1;X,T_1}, \dots, A^{P_d(\mathrm{m}_d)}_{M_d;X,T_d}$ commute. Thus it remains to verify (4.2) and (4.3). We fix $j \in [d]$. For (4.2), we refer to Theorem 1.4 in [54], which ensures that for every $p \in (1, \infty)$, one has

$$\sup_{J\in\mathbb{Z}_+} \sup_{I\in\mathfrak{S}_J(\mathbb{Z}_+)} \|O^2_{I,J}(A^{P_j(\mathrm{m}_j)}_{M_j;X,T_j} : M_j \in \mathbb{Z}_+)\|_{L^p(X)} \lesssim_p \|f\|_{L^p(X)}, \quad f \in L^p(X).$$

For (4.3) we refer to Theorem C in [56], which guarantees that for every $p \in (1, \infty)$ one has

$$\Big\| \Big(\sum_{\iota\in\mathbb{Z}} \big(\sup_{M_j\in\mathbb{Z}_+} |A^{P_j(\mathrm{m}_j)}_{M_j;X,T_j} |f_\iota|\big)^2 \Big)^{1/2} \Big\|_{L^p(X)} \lesssim_p \Big\| \Big(\sum_{\iota\in\mathbb{Z}} |f_\iota|^2 \Big)^{1/2} \Big\|_{L^p(X)},$$

uniformly in $(f_j)_{j\in\mathbb{Z}} \in L^p(X; \ell^2(\mathbb{Z}))$. This completes the proof of the multi-parameter oscillation inequality (1.27) in Theorem 1.25. ∎

Funding. Mariusz Mirek was partially supported by NSF grant DMS-2154712, and by the National Science Centre in Poland, grant Opus 2018/31/B/ST1/00204. Tomasz Z. Szarek was partially supported by the National Science Centre of Poland, grant Opus 2017/27/B/ST1/01623, by Juan de la Cierva Incorporación 2019 grant number IJC2019-039661-I funded by Agencia Estatal de Investigación, grant PID2020-113156GB-I00/AEI/10.13039/501100011033, and also by the Basque Government through the BERC 2022-2025 program and by Spanish Ministry of Sciences, Innovation and Universities: BCAM Severo Ochoa accreditation SEV-2017-0718.

References

[1] Bellow, A.: Two problems submitted by A. Bellow. In *Measure theory Oberwolfach 1981 (Proceedings of the Conference held at Oberwolfach, June 21–27, 1981)*, pp 429–431. Lecture Notes in Mathematics 945, Springer-Verlag, Berlin-Heidelberg, 1982.

[2] Beltran, D., Oberlin, R., Roncal, L., Seeger, A. and Stovall, B.: Variation bounds for spherical averages. *Math. Ann.* **382** (2022), no. 1-2, 459–512.

[3] Beltran, D., Roos, J. and Seeger, A.: Multi-scale sparse domination. To appear in *Mem. Amer. Math. Soc.*

[4] Bergelson, V. and Leibman, A.: A nilpotent Roth theorem. *Invent. Math.* **147** (2002), no. 2, 429–470.

[5] Birkhoff, G. D.: Proof of the ergodic theorem. *Proc. Natl. Acad. Sci. USA* **17** (1931), no. 12, 656–660.

[6] Bourgain, J.: On the maximal ergodic theorem for certain subsets of the integers. *Israel J. Math.* **61** (1988), no. 1, 39–72.

[7] Bourgain, J.: On the pointwise ergodic theorem on L^p for arithmetic sets. *Israel J. Math.* **61** (1988), no. 1, 73–84.

[8] Bourgain, J.: Pointwise ergodic theorems for arithmetic sets. (With an appendix by the author, H. Furstenberg, Y. Katznelson, and D. S. Ornstein.) *Inst. Hautes Études Sci. Publ. Math.* **69** (1989), 5–45.

[9] Bourgain, J., Mirek, M., Stein, E. M. and Wright, J.: On a multi-parameter variant of the Bellow–Furstenberg problem. Preprint 2022, arXiv: 2209.07358.

[10] Buczolich, Z. and Mauldin, R. D.: Divergent square averages. *Ann. of Math. (2)* **171** (2010), no. 3, 1479–1530.

[11] Burkholder, D. L.: Martingale transforms. *Ann. Math. Statist.* **37** (1966), 1494–1504.

[12] Burkholder, D. L.: Explorations in martingale theory and its applications. In *École d'Été de Probabilités de Saint-Flour XIX-1989*, pp. 1–66. Lecture Notes in Math. 1464, Springer, Berlin, 1991.

[13] Calderón, A. P.: Ergodic theory and translation invariant operators. *Proc. Natl. Acad. Sci. USA* **59** (1968), 349–353.

[14] Campbell, J. T., Jones, R. L., Reinhold, K. and Wierdl, M.: Oscillation and variation for the Hilbert transform. *Duke Math. J.* **105** (2000), no. 1, 59–83.

[15] Carleson, L.: On convergence and growth of partial sums of Fourier series. *Acta Math.* **116** (1966), 135–157.

[16] Cowling, M. and Tao, T.: Some light on Littlewood–Paley theory. *Math. Ann.* **321** (2001), no. 4, 885–888.

[17] Di Plinio, F., Do, Y. and Uraltsev, G.: Positive sparse domination of variational Carleson operators. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* **18** (2018), no. 4, 1443–1458.

[18] Do, Y. and Lacey, M.: Weighted bounds for variational Fourier series. *Studia Math.* **211** (2012), no. 2, 153–190.

[19] Dunford, N.: An individual ergodic theorem for non-commutative transformations. *Acta Sci. Math. (Szeged)* **14** (1951), 1–4.

[20] Duoandikoetxea, J. and Rubio de Francia, J. L.: Maximal and singular integral operators via Fourier transform estimates. *Invent. Math.* **84** (1986), no. 3, 541–561.

[21] Einsiedler, M. and Ward, T.: *Ergodic theory with a view towards number theory*. Graduate Texts in Mathematics 259, Springer-Verlag, London, 2011.

[22] Fefferman, C.: Pointwise convergence of Fourier series. *Ann. of Math. (2)* **98** (1973), 551–571.

[23] Furstenberg, H.: *Problems session, Conference on ergodic theory and applications*. University of New Hampshire, Durham, NH, June 1982.

[24] Grafakos, L., Martell, J. M. and Soria, F.: Weighted norm inequalities for maximally modulated singular integral operators. *Math. Ann.* **331** (2005), no. 2, 359–394.

[25] Grafakos, L., Tao, T. and Terwilleger, E.: L^p bounds for a maximal dyadic sum operator. *Math. Z.* **246** (2004), no. 1-2, 321–337.

[26] Guo, S., Roos, J. and Yung, P.-L.: Sharp variation-norm estimates for oscillatory integrals related to Carleson's theorem. *Anal. PDE* **13** (2020), no. 5, 1457–1500.

[27] Hunt, R.: On the convergence of Fourier series. In *Orthogonal expansions and their continuous analogues (Proc. Conf. Edwardsville, IL, 1967)*, pp. 235–255. Southern Illinois Univ. Press, Carbondale, IL, 1968.

[28] Hytönen, T., van Neerven, J., Veraar, M. and Weis, L.: *Analysis in Banach spaces. Vol. I. Martingales and Littlewood–Paley theory*. A Series of Modern Surveys in Mathematics 63, Springer, Cham, 2016.

[29] Ionescu, A. D., Magyar, Á., Mirek, M. and Szarek, T. Z.: Polynomial averages and pointwise ergodic theorems on nilpotent groups. To appear in *Invent. Math.*, DOI s00222-022-01159-0.

[30] Ionescu, A., Magyar, Á., Stein, E. M. and Wainger, S.: Discrete Radon transforms and applications to ergodic theory. *Acta Math.* **198** (2007), no. 2, 231–298.

[31] Ionescu, A., Magyar, Á. and Wainger, S.: Averages along polynomial sequences in discrete nilpotent Lie groups: Singular Radon transforms. In *Advances in analysis: the legacy of Elias M. Stein*, pp. 146–188. Princeton Math. Ser. 50, Princeton Univ. Press, Princeton, NJ, 2014.

[32] Ionescu, A. D. and Wainger, S.: L^p boundedness of discrete singular Radon transforms. *J. Amer. Math. Soc.* **19** (2005), no. 2, 357–383.

[33] Jones, R. L., Kaufman, R., Rosenblatt, J. M. and Wierdl, M.: Oscillation in ergodic theory. *Ergodic Theory Dynam. Systems* **18** (1998), no. 4, 889–935.

[34] Jones, R. L. and Reinhold, K.: Oscillation and variation inequalities for convolution powers. *Ergodic Theory Dynam. Systems* **21** (2001), no. 6, 1809–1829.

[35] Jones, R. L., Rosenblatt, J. M. and Wierdl, M.: Oscillation inequalities for rectangles. *Proc. Amer. Math. Soc.* **129** (2001), no. 5, 1349–1358.

[36] Jones, R. L., Rosenblatt, J. M. and Wierdl, M.: Oscillation in ergodic theory: higher dimensional results. *Israel J. Math.* **135** (2003), 1–27.

[37] Jones, R. L., Seeger, A. and Wright, J.: Strong variational and jump inequalities in harmonic analysis. *Trans. Amer. Math. Soc.* **360** (2008), no. 12, 6711–6742.

[38] Jones, R. L. and Wang, G.: Variation inequalities for the Fejér and Poisson kernels. *Trans. Amer. Math. Soc.* **356** (2004), no. 11, 4493–4518.

[39] Kenig, C. E. and Tomas, P. A.: Maximal operators defined by Fourier multipliers. *Studia Math.* **68** (1980), no. 1, 79–83.

[40] Kolmogorov, A. N.: Une série de Fourier–Lebesgue divergente presque partout. *Fund. Math.* **4** (1923), no. 1, 324–328.

[41] Kolmogorov, A. N.: Une série de Fourier–Lebesgue divergente partout. *C. R. Acad. Sci. Paris* **183** (1926), 1327–1329.

[42] Krause, B.: Polynomial ergodic averages converge rapidly: variations on a theorem of Bourgain. To appear in *Israel J. Math.*

[43] Krause, B., Mirek, M. and Tao, T.: Pointwise ergodic theorems for non-conventional bilinear polynomial averages. *Ann. of Math. (2)* **195** (2022), no. 3, 997–1109.

[44] Lacey, M.: Personal communication.

[45] Lacey, M.: Carleson's theorem: proof, complements, variations. *Publ. Mat.* **48** (2004), no. 2, 251–307.

[46] Lacey, M.: Sparse bounds for spherical maximal functions. *J. Anal. Math.* **139** (2019), no. 2, 613–635.

[47] Lacey, M. and Terwilleger, E.: A Wiener–Wintner theorem for the Hilbert transform. *Ark. Mat.* **46** (2008), no. 2, 315–336.

[48] Lacey, M. and Thiele, C.: A proof of boundedness of the Carleson operator. *Math. Res. Lett.* **7** (2000), no. 4, 361–370.

[49] LaVictoire, P.: Universally L^1-bad arithmetic sequences. *J. Anal. Math.* **113** (2011), no. 1, 241–263.

[50] Lépingle, D.: La variation d'ordre p des semi-martingales. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.* **36** (1976), no. 4, 295–316.

[51] Magyar, Á, Stein, E. M. and Wainger, S.: Discrete analogues in harmonic analysis: spherical averages. *Ann. of Math. (2)* **155** (2002), no. 1, 189–208.

[52] Menshov, D.: Sur les séries de fonctions orthogonales. *Fund. Math.* **4** (1923), 82–105.

[53] Mirek, M.: $\ell^p(\mathbb{Z}^d)$-estimates for discrete Radon transform: square function estimates. *Anal. PDE* **11** (2018), no. 3, 583–608.

[54] Mirek, M., Słomian, W. and Szarek, T. Z.: Some remarks on oscillation inequalities. To appear in *Ergodic Theory Dynam. Systems*, DOI 10.1017/etds.2022.77.

[55] Mirek, M., Stein, E. M. and Trojan, B.: $\ell^p(\mathbb{Z}^d)$-estimates for discrete operators of Radon type: Variational estimates. *Invent. Math.* **209** (2017), no. 3, 665–748.

[56] Mirek, M., Stein, E. M. and Trojan, B.: $\ell^p(\mathbb{Z}^d)$-estimates for discrete operators of Radon type: Maximal functions and vector-valued estimates. *J. Funct. Anal*. **277** (2019), no. 8, 2471–2521.

[57] Mirek, M., Stein, E. M. and Zorin-Kranich, P.: Jump inequalities via real interpolation. *Math. Ann.* **376** (2020), no. 1-2, 797–819.

[58] Mirek, M., Stein, E. M. and Zorin-Kranich, P.: A bootstrapping approach to jump inequalities and their applications. *Anal. PDE* **13** (2020), no. 2, 527–558.

[59] Mirek, M., Stein, E. M. and Zorin-Kranich, P.: Jump inequalities for translation-invariant operators of Radon type on $\mathbb{Z}^d$. *Adv. Math.* **365** (2020), Paper no. 107065, 57 pp.

[60] Mirek, M. and Trojan, B.: Discrete maximal functions in higher dimensions and applications to ergodic theory. *Amer. J. Math.* **138** (2016), no. 6, 1495–1532.

[61] von Neumann, J.: Proof of the quasi-ergodic hypothesis. *Proc. Natl. Acad. Sci. USA* **18** (1932), 70–82.

[62] Nevo, A.: Pointwise ergodic theorems for actions of groups. In *Handbook of dynamical systems, Volume 1B*, pp. 871–982. Elsevier Science, Amsterdam, 2005.

[63] Oberlin, R., Seeger, A., Tao, T., Thiele, C. and Wright, J.: A variation norm Carleson theorem. *J. Eur. Math. Soc. (JEMS)* **14** (2012), no. 2, 421–464.

[64] Petersen, K.: *Ergodic theory*. Cambridge Studies in Advanced Mathematics 2, Cambridge University Press, 1989.

[65] Pierce, L. B.: On superorthogonality. *J. Geom. Anal*. **31** (2021), no. 7, 7096–7183.

[66] Pisier, G. and Xu, Q. H.: The strong p-variation of martingales and orthogonal series. *Probab. Theory Related Fields* **77** (1988), no. 4, 497–514.

[67] Pramanik, M. and Terwilleger, E.: A weak L^2 estimate for a maximal dyadic sum operator on $\mathbb{R}^n$. *Illinois J. Math.* **47** (2003), no. 3, 775–813.

[68] Rademacher, H.: Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen. *Math. Ann.* **87** (1922), no. 1-2, 112–138.

[69] Riesz, F.: Some mean ergodic theorems. *J. London Math. Soc.* **13** (1938), no. 4, 274–278.

[70] Rosenblatt, J. and Wierdl, M.: Pointwise ergodic theorems via harmonic analysis. In *Ergodic theory and its connections with harmonic analysis (Alexandria, 1993)*, pp. 3–151. London Mathematical Society Lecture Notes 205, Cambridge Univ. Press, Cambridge, 1995.

[71] Rubio de Francia, J. L.: A Littlewood–Paley inequality for arbitrary intervals. *Rev. Mat. Iberoamericana* **1** (1985), no. 2, 1–14.

[72] Stein, E. M.: *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*. Princeton Mathematical Series 43, Monographs in Harmonic Analysis III, Princeton University Press, Princeton, NJ, 1993.

[73] Szemerédi, E.: On sets of integers containing no k elements in arithmetic progression. *Acta Arith.* **27** (1975), 199–245.

[74] Tao, T.: The Ionescu–Wainger multiplier theorem and the adeles. *Mathematika* **67** (2021), no. 3, 647–677.

[75] Uraltsev, G.: Variational Carleson embeddings into the upper 3-space. Preprint 2016, arXiv: 1610.07657.

[76] Zorin–Kranich, P.: Variation estimates for averages along primes and polynomials. *J. Funct. Anal.* **268** (2015), no. 1, 210–238.

[77] Zygmund, A.: An individual ergodic theorem for non-commutative transformations. *Acta Sci. Math. (Szeged)* **14** (1951), 103–110.

Received November 25, 2021; revised September 2, 2022. Published online December 23, 2022.

Mariusz Mirek
Department of Mathematics, Rutgers University, Piscataway, NJ 08854-8019;
and School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA;
and Instytut Matematyczny, Uniwersytet Wrocławski, Plac Grunwaldzki 2,
50-384 Wrocław, Poland;
mariusz.mirek@rutgers.edu

Tomasz Z. Szarek
BCAM-Basque Center for Applied Mathematics, 48009 Bilbao, Spain;
and Instytut Matematyczny, Uniwersytet Wrocławski, Plac Grunwaldzki 2,
50-384 Wrocław, Poland;
tzszarek@bcamath.org

James Wright
Maxwell Institute of Mathematical Sciences and The School of Mathematics,
The University of Edinburgh, James Clerk Maxwell Building, The King's Buildings,
Peter Guthrie Tait Road, City Edinburgh EH9 3FD, UK;
J.R.Wright@ed.ac.uk