CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

On a multi-parameter variant of the Bellow–Furstenberg problem

Jean Bourgain[1], Mariusz Mirek [2], Elias M. Stein[3] and James Wright[4]

[1]School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA; E-mail: bourgain@math.ias.edu.
[2]Department of Mathematics, Rutgers University, Piscataway, NJ 08854-8019, USA & School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA & Instytut Matematyczny, Uniwersytet Wrocławski, Plac Grunwaldzki 2/4, 50-384 Wrocław Poland; E-mail: mariusz.mirek@rutgers.edu.
[3]Department of Mathematics, Princeton University, Princeton, NJ 08544-100, USA; E-mail: stein@math.princeton.edu.
[4]James Clerk Maxwell Building, The King's Buildings, Peter Guthrie Tait Road, City Edinburgh, EH9 3FD;
E-mail: J.R.Wright@ed.ac.uk.

Abstract

We prove convergence in norm and pointwise almost everywhere on L^p, $p \in (1, \infty)$, for certain multi-parameter polynomial ergodic averages by establishing the corresponding multi-parameter maximal and oscillation inequalities. Our result, in particular, gives an affirmative answer to a multi-parameter variant of the Bellow–Furstenberg problem. This paper is also the first systematic treatment of multi-parameter oscillation semi-norms which allows an efficient handling of multi-parameter pointwise convergence problems with arithmetic features. The methods of proof of our main result develop estimates for multi-parameter exponential sums, as well as introduce new ideas from the so-called multi-parameter circle method in the context of the geometry of backwards Newton diagrams that are dictated by the shape of the polynomials defining our ergodic averages.

Contents

1. Introduction

1.1. A brief history

In 1933, Khintchin [40] had the great insight to see how to generalize the classical equidistribution result of Bohl [12], Sierpiński [56] and Weyl [66] from 1910 to a pointwise ergodic theorem, observing that as a consequence of Birkhoff's famous ergodic theorem [11], the following equidistribution result holds: namely, for any irrational $\theta \in \mathbb{R}$, for any Lebesgue measurable set $E \subseteq [0, 1)$ and for almost every $x \in \mathbb{R}$,

$$\lim_{M \to \infty} \frac{\#\{m \in [M] : \{x + m\theta\} \in E\}}{M} = |E|,$$

where $\{x\}$ denotes the fractional part of $x \in \mathbb{R}$, and $[N] := (0, N] \cap \mathbb{Z}$ for any real number $N \geq 1$. In 1916, Weyl [67] extended the classical equidistribution theorem to general polynomial sequences $(\{P(n)\})_{n \in \mathbb{N}}$ having at least one irrational coefficient, and so it was natural to ask whether a pointwise ergodic extension of Weyl's equidistribution theorem holds. This question was posed by Bellow [5] and Furstenberg [24] in the early 1980s; precisely, they asked if for any polynomial $P \in \mathbb{Z}[m]$ with integer coefficients and $P(0) = 0$ and for any invertible measure-preserving transformation $T : X \to X$ on a probability space $(X, \mathcal{B}(X), \mu)$, does the limit

$$\lim_{M \to \infty} \mathbb{E}_{m \in [M]} f(T^{P(m)}x)$$

exist for almost every $x \in X$ and for every $f \in L^\infty(X)$? Here and throughout the paper we use the notation $\mathbb{E}_{y \in Y} f(y) := \frac{1}{\#Y} \sum_{y \in Y} f(y)$ for any finite set $Y \neq \emptyset$ and any function $f : Y \to \mathbb{C}$. In the mid 1980s, the first author [13, 14, 15] established that this is indeed the case whenever $f \in L^p(X)$ and $p \in (1, \infty)$, leaving open the question of what happens on $L^1(X)$. Interestingly, it was shown much later by Buczolich and Mauldin [18] that the above pointwise convergence result fails for general L^1 functions when $P(m) = m^2$; see also [42] for further refinements. In any case, the papers [13, 14, 15] represent a far-reaching common generalization of Birkhoff's pointwise ergodic theorem and Weyl's equidistribution theorem.

Both Birkhoff's and Weyl's results have natural multi-parameter extensions. In 1951, Dunford [23] and Zygmund [72] independently extended Birkhoff's theorem to multiple measure-preserving transformations $T_1, \ldots, T_k : X \to X$. They showed that the limit

$$\lim_{M_1, \ldots, M_k \to \infty} \mathbb{E}_{(m_1, \ldots, m_k) \in \prod_{j=1}^{k} [M_j]} f(T_1^{m_1} \cdots T_k^{m_k} x) \tag{1.1}$$

exists for almost every $x \in X$ and for any $f \in L^p(X)$ with $p \in (1, \infty)$, where $\prod_{j=1}^{k} [M_j] := [M_1] \times \ldots \times [M_k]$. The limit is taken in the unrestricted sense; that is, when $\min\{M_1, \ldots, M_k\} \to \infty$. Here, when $k \geq 2$, the pointwise convergence result is manifestly false for general $f \in L^1(X)$.

In 1979, Arkhipov, Chubarikov and Karatsuba [2] extended Weyl's equidistribution result to polynomials (even multiple polynomials) of several variables. In its simplest form, their result asserts that for any k-variate polynomial $P \in \mathbb{Z}[m_1, \ldots, m_k]$, any irrational $\theta \in \mathbb{R}$ and any interval $[a, b) \subseteq [0, 1)$, one has

$$\lim_{\min\{M_1, \ldots, M_k\} \to \infty} \frac{\#\{(m_1, \ldots, m_k) \in \prod_{j=1}^{k} [M_j] : \{\theta P(m_1, \ldots, m_k)\} \in [a, b)\}}{M_1 \cdots M_k} = b - a. \tag{1.2}$$

In the late 1980s, after [13, 14, 15] and in light of these results, it was natural to seek a common generalization of the results of Dunford and Zygmund on the one hand (which generalize Birkhoff's original theorem) and Arkhipov, Chubarikov and Karatsuba on the other hand (which generalize Weyl's theorem), which can be subsumed under the following conjecture, a multi-parameter variant of the Bellow–Furstenberg problem:

Conjecture 1.3. *Let $k \in \mathbb{Z}_+$ with $k \geq 2$ be given and let $(X, \mathcal{B}(X), \mu)$ be a probability measure space with an invertible measure-preserving transformation $T : X \to X$. Assume that $P \in \mathbb{Z}[m_1, \ldots, m_k]$ with $P(0) = 0$. Then for any $f \in L^\infty(X)$, the limit*

$$\lim_{\min\{M_1, \ldots, M_k\} \to \infty} \mathbb{E}_{(m_1, \ldots, m_k) \in \prod_{j=1}^{k} [M_j]} f(T^{P(m_1, \ldots, m_k)} x) \quad \text{exists for } \mu\text{-almost every } x \in X. \tag{1.4}$$

Our main theorem resolves this conjecture.

Theorem 1.5. *Conjecture 1.3 is true for all $k \in \mathbb{Z}_+$.*

The case $k = 1$ corresponds to the classical one-parameter question of Bellow [5] and Furstenberg [24] and was resolved in [13, 14, 15]. In this paper, we will establish the cases $k \geq 2$. In fact, we will prove stronger quantitative results including corresponding multi-parameter maximal and oscillation estimates (see Theorem 1.11 below), which will imply Conjecture 1.3. This paper also represents a first systematic treatment of multi-parameter oscillation semi-norms which allows an efficient handling of multi-parameter pointwise convergence problems for ergodic averaging operators with polynomial orbits. Before we formulate our main quantitative results, we briefly describe the interesting history of Conjecture 1.3.

The theorems of Dunford [23] and Zygmund [72] have simple proofs, which can be deduced by iterative applications of the classical Birkhoff ergodic theorem. For this purpose, it suffices to note that the Dunford–Zygmund averages from (1.1) can be written as a composition of k classical Birkhoff averages as follows

$$\mathbb{E}_{(m_1,\ldots,m_k) \in \prod_{j=1}^{k}[M_j]} f(T_1^{m_1} \cdots T_k^{m_k} x) = \mathbb{E}_{m_k \in [M_k]} \big[\cdots \mathbb{E}_{m_1 \in [M_1]} f(T_1^{m_1}(\cdots T_k^{m_k})x) \big]. \qquad (1.6)$$

The order in this composition is important since the transformations $T_1, \ldots, T_k$ do not need to commute. The first author, in view of [13, 14, 15], extended the observation from (1.6) to polynomial orbits and showed that for every $f \in L^p(X)$ with $p \in (1, \infty)$, the limit

$$\lim_{\min\{M_1,\ldots,M_k\} \to \infty} \mathbb{E}_{(m_1,\ldots,m_k) \in \prod_{j=1}^{k}[M_j]} f(T_1^{P_1(m_1)} \cdots T_k^{P_k(m_k)} x) \qquad (1.7)$$

exists for μ-almost every $x \in X$, whenever $P_1, \ldots, P_k \in \mathbb{Z}[\mathrm{m}]$ with $P_1(0) = \ldots = P_k(0) = 0$ and $T_1, \ldots, T_d : X \to X$ is a family of commuting and invertible measure-preserving transformations. The result from (1.7) was never published; nonetheless, it can be thought of as a polynomial extension of the theorem of Dunford [23] and Zygmund [72] (the arguments in Section 3.4 can be used to derive a quantitative version of (1.7)). Interestingly, as observed by Benjamin Weiss (privately communicated to the first author), any ergodic theorem for these averages fails in general for $k \geq 2$ when the $T_1, \ldots, T_k$ are general non-commuting transformations. It may even fail in the one-parameter situation for the averages of the form $\mathbb{E}_{m \in [M]} f(T_1^{P_1(m)} \cdots T_k^{P_k(m)} x)$; see also [10] for interesting counterexamples.

This was a turning point, illustrating that the multi-parameter theory for averages with orbits along polynomials with separated variables as in (1.7) is well-understood and can be readily deduced from the one-parameter theory [13, 14, 15] by simple iteration as in (1.6). However, the equidistribution result (1.2) of Arkhipov, Chubarikov and Karatsuba [2], based on the so-called multi-parameter circle method (deep and intricate tools in analytic number theory which go beyond the classical circle method) showed that the situation may be dramatically different when orbits are defined along genuinely k-variate polynomials $P \in \mathbb{Z}[m_1, \ldots, m_k]$ and led to Conjecture 1.3. Even for $k = 2$ with $P(m_1, m_2) = m_1^2 m_2^3$ in (1.4), the problem becomes very challenging. Surprisingly it seems that there is no simple way (like changing variables or interpreting the average from (1.4) as a composition of simpler one-parameter averages as in (1.6)) that would help us to reduce the matter to the setup where pointwise convergence is known.

The multi-parameter case $k \geq 2$ in Conjecture 1.3 lies in sharp contrast to the one-parameter situation $k = 1$, causing serious difficulties that were not apparent in [13, 14, 15]. The most notable differences are multi-parameter estimates of corresponding exponential sums and a delicate control of error terms that arise in implementing the circle method. These difficulties arise from the lack of nestedness when the parameters $M_1, \ldots, M_k$ are independent; see Figure 1 and Figure 2 below. We now turn to a more detailed discussion and precise formulation of the results in this paper.
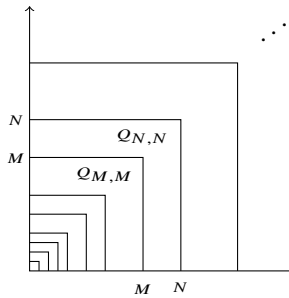
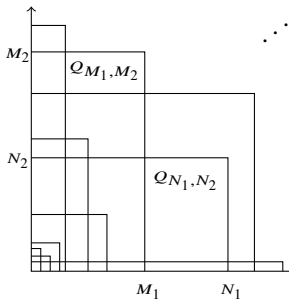**Figure 1.** *Family of nested rectangles (cubes) $Q_{M,M} \subset Q_{N,N}$ with $M < N$, for $k = 2$.*

Figure 2. *Family of un-nested rectangles $Q_{M_1,M_2} \nsubseteq Q_{N_1,N_2}$ with $M_1 < N_1$ and $M_2 > N_2$, for $k = 2$.*

1.2. Statement of the main results

Throughout this paper, the triple $(X, \mathcal{B}(X), \mu)$ denotes a σ-finite measure space, and $\mathbb{Z}[m_1, \ldots, m_k]$ denotes the space of all formal k-variate polynomials $P(m_1, \ldots, m_k)$ with $k \in \mathbb{Z}_+$ indeterminates $m_1, \ldots, m_k$ and integer coefficients. Each polynomial $P \in \mathbb{Z}[m_1, \ldots, m_k]$ will always be identified with a map $\mathbb{Z}^k \ni (m_1, \ldots, m_k) \mapsto P(m_1, \ldots, m_k) \in \mathbb{Z}$.

Let $d, k \in \mathbb{Z}_+$, and given a family $\mathcal{T} = \{T_1, \ldots, T_d\}$ of invertible commuting measure-preserving transformations on X, a measurable function f on X, polynomials $\mathcal{P} = \{P_1, \ldots, P_d\} \subset \mathbb{Z}[m_1, \ldots, m_k]$ and a vector of real numbers $M = (M_1, \ldots, M_k)$ whose entries are greater than 1, we define the multi-parameter polynomial ergodic average by

$$A^{\mathcal{P}}_{M;X,\mathcal{T}} f(x) := \mathbb{E}_{m \in Q_M} f(T_1^{P_1(m)} \cdots T_d^{P_d(m)} x), \qquad x \in X, \tag{1.8}$$

where $Q_M := [M_1] \times \ldots \times [M_k]$ is a rectangle in $\mathbb{Z}^k$. We will often abbreviate $A^{\mathcal{P}}_{M;X,\mathcal{T}}$ to $A^{\mathcal{P}}_{M;X}$ when the tranformations are understood. In some instances, we will write out the averages

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

depending on how explicit we want to be.

Example 1.9. From the point of view of pointwise convergence problems, due to the Calderón transference principle [19], the most important dynamical system is the integer shift system. Consider the d-dimensional lattice $(\mathbb{Z}^d, \mathcal{B}(\mathbb{Z}^d), \mu_{\mathbb{Z}^d})$ equipped with a family of shifts $S_1, \ldots, S_d : \mathbb{Z}^d \to \mathbb{Z}^d$, 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 j-th basis vector from the standard basis in $\mathbb{Z}^d$ for

each $j \in [d]$). The average $A^{\mathcal{P}}_{M;X,\mathcal{T}}$ with $\mathcal{T} = (T_1, \ldots, T_d) = (S_1, \ldots, S_d)$ can be rewritten for any $x = (x_1, \ldots, x_d) \in \mathbb{Z}^d$ and any finitely supported function $f : \mathbb{Z}^d \to \mathbb{C}$ as

$$A^{\mathcal{P}}_{M;\mathbb{Z}^d} f(x) = \mathbb{E}_{m \in Q_M} f(x_1 - P_1(m), \ldots, x_d - P_d(m)). \tag{1.10}$$

The main result of this paper, which implies Conjecture 1.3, is the following ergodic theorem.

Theorem 1.11. *Let $(X, \mathcal{B}(X), \mu)$ be a σ-finite measure space with an invertible measure-preserving transformation $T : X \to X$. Let $k \in \mathbb{Z}_+$ with $k \geq 2$ be given, and $P \in \mathbb{Z}[\mathrm{m}_1, \ldots, \mathrm{m}_k]$ be a polynomial such that $P(0) = 0$. Let $f \in L^p(X)$ for some $1 \leq p \leq \infty$, and let $A^P_{M_1, \ldots, M_k; X, T} f$ be the average defined in (1.8) with $d = 1$ and arbitrary $k \in \mathbb{Z}_+$.*

(i) *(Mean ergodic theorem) If $1 < p < \infty$, then the averages $A^P_{M_1, \ldots, M_k; X, T} f$ converge in $L^p(X)$ norm.*

(ii) *(Pointwise ergodic theorem) If $1 < p < \infty$, then the averages $A^P_{M_1, \ldots, M_k; X, T} f$ converge pointwise almost everywhere.*

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

$$\left\| \sup_{M_1, \ldots, M_k \in \mathbb{Z}_+} |A^P_{M_1, \ldots, M_k; X, T} f| \right\|_{L^p(X)} \lesssim_{p,P} \|f\|_{L^p(X)}. \tag{1.12}$$

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

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^k)} \left\| O_{I,J} (A^P_{M_1, \ldots, M_k; X, T} f : M_1, \ldots, M_k \in \mathbb{D}_\tau) \right\|_{L^p(X)} \lesssim_{p,\tau,P} \|f\|_{L^p(X)}, \tag{1.13}$$

where $\mathbb{D}_\tau := \{\tau^n : n \in \mathbb{N}\}$; see Section 2 for a definition of the oscillation semi-norm $O_{I,J}$. The implicit constant in (1.12) and (1.13) may depend on p, τ, P.

For ease of exposition, we only prove Theorem 1.11 in the two-parameter setting $k = 2$, though there are some places in the paper where some arguments are formulated and proved in the multi-parameter setting to convince the reader that our arguments are adaptable to the general multi-parameter setup. However, the patient reader will readily see that all two-parameter arguments are adaptable (at the expense of introducing cumbersome notation, which would make the exposition unreadable) to the general multi-parameter setting for arbitrary $k \geq 2$, by multiple iterations of the arguments presented in the paper.

We now give some remarks about Theorem 1.11.

1. Theorem 1.11 establishes Conjecture 1.3 for the averages $A^P_{M;X,T} f$. This is the first nontrivial result in the literature establishing pointwise almost everywhere convergence for polynomial ergodic averages in the multi-parameter setting. See [53] for other pointwise convergence results in the multi-parameter setting.

2. The proof of Theorem 1.11 is relatively simple if $P \in \mathbb{Z}[\mathrm{m}_1, \ldots, \mathrm{m}_k]$ is degenerate; see inequality (3.5) in Section 3. We will say that $P \in \mathbb{Z}[\mathrm{m}_1, \ldots, \mathrm{m}_k]$ is degenerate if it can be written as

$$P(\mathrm{m}_1, \ldots, \mathrm{m}_k) = P_1(\mathrm{m}_1) + \ldots + P_k(\mathrm{m}_k), \tag{1.14}$$

where $P_1 \in \mathbb{Z}[\mathrm{m}_1], \ldots, P_k \in \mathbb{Z}[\mathrm{m}_k]$ with $P_1(0) = \ldots = P_k(0) = 0$. Otherwise, we say that $P \in \mathbb{Z}[\mathrm{m}_1, \ldots, \mathrm{m}_k]$ is non-degenerate. The method of proof of Theorem 1.11 in the degenerate case can be also used to derive quantitative oscillation bounds for the polynomial Dunford and Zygmund theorem establishing (1.7).

3. At the expense of great complexity, one can also prove that inequality (1.13) holds with $\mathbb{Z}_+$ in place of $\mathbb{D}_\tau$. However, we do not address this question here, since (1.13) is sufficient for our purposes and will allow us to establish Theorem 1.11(ii).

4. If $(X, \mathcal{B}(X), \mu)$ is a probability space and the measure preserving transformation T in Theorem 1.11 is totally ergodic, then Theorem 1.11(ii) implies

$$\lim_{\min\{M_1,\ldots,M_k\} \to \infty} A^P_{M_1,\ldots,M_k;X,T} f(x) = \int_X f(y) d\mu(y) \qquad (1.15)$$

μ-almost everywhere on X. We recall that a measure preserving transformation T is called *ergodic* on X if $T^{-1}[B] = B$ implies $\mu(B) = 0$ or $\mu(B) = 1$, and *totally ergodic* if T^n is ergodic for every $n \in \mathbb{Z}_+$.

5. This paper is the first systematic treatment of multi-parameter oscillation semi-norms; see (2.9), Proposition 2.16 and Proposition 2.18. Moreover, it seems that the 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 jumps to handle pointwise convergence problems; see [38, 49] and the references therein. Multi-parameter oscillations (2.9) were considered for the first time in [37] in the context of the Dunford–Zygmund averages (1.1) for commuting measure-preserving transformations.

We close this subsection by emphasizing that the methods developed in this paper allow us to handle averages (1.8) with multiple polynomials. At the expense of some additional work, one can prove the following ergodic theorem.

Theorem 1.16. *Let $(X, \mathcal{B}(X), \mu)$ be a σ-finite measure space equipped with a family of commuting invertible and measure-preserving transformations $T_1, T_2, T_3 : X \to X$. Let $P \in \mathbb{Z}[m_1, m_2]$ be a polynomial such that $P(0, 0) = \partial_1 P(0, 0) = \partial_2 P(0, 0) = 0$, which additionally has partial degrees (as a polynomial of the variable m_1 and a polynomial of the variable m_2) at least two. Let $f \in L^p(X)$ for some $1 \le p \le \infty$, and let $A^{m_1,m_2,P(m_1,m_2)}_{M_1,M_2;X} f$ be the average defined in (1.8) with $d = 3$, $k = 2$, and $P_1(m_1, m_2) = m_1$, $P_2(m_1, m_2) = m_2$ and $P_3(m_1, m_2) = P(m_1, m_2)$.*

(i) *(Mean ergodic theorem) If $1 < p < \infty$, then the averages $A^{m_1,m_2,P(m_1,m_2)}_{M_1,M_2;X} f$ converge in $L^p(X)$ norm.*

(ii) *(Pointwise ergodic theorem) If $1 < p < \infty$, then the averages $A^{m_1,m_2,P(m_1,m_2)}_{M_1,M_2;X} f$ converge pointwise almost everywhere.*

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

$$\left\| \sup_{M_1,M_2 \in \mathbb{Z}_+} |A^{m_1,m_2,P(m_1,m_2)}_{M_1,M_2;X} f| \right\|_{L^p(X)} \lesssim_{p,P} \|f\|_{L^p(X)}. \qquad (1.17)$$

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

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^2)} \left\| O_{I,J} (A^{m_1,m_2,P(m_1,m_2)}_{M_1,M_2;X} f : M_1, M_2 \in \mathbb{D}_\tau) \right\|_{L^p(X)} \lesssim_{p,\tau,P} \|f\|_{L^p(X)},$$

$$(1.18)$$

where $\mathbb{D}_\tau := \{\tau^n : n \in \mathbb{N}\}$. The implicit constant in (1.17) and (1.18) may depend on p, τ, P.

For simplicity of notation, we have only formulated Theorem 1.16 in the two-parameter setting, but it can be extended to a multi-parameter setting as well. Namely, let $d \ge 2$ and let $(X, \mathcal{B}(X), \mu)$ be a σ-finite measure space equipped with a family of commuting invertible and measure-preserving transformations $T_1, \ldots, T_d : X \to X$. Suppose that $P \in \mathbb{Z}[m_1, \ldots, m_{d-1}]$ is a polynomial such that

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

which has partial degrees (as a polynomial of the variable m_i for any $i \in [d-1]$) at least two. Then the conclusions of Theorem 1.16 remain true for the averages

$$A^{m_1,\ldots,m_{d-1},P(m_1,\ldots,m_{d-1})}_{M_1,\ldots,M_{d-1};X,T_1,\ldots,T_d} f \quad \text{in place of} \quad A^{m_1,m_2,P(m_1,m_2)}_{M_1,M_2;X} f. \tag{1.19}$$

All remarks from items 1–4 after Theorem 1.11 remain true for ergodic averages from (1.19). Finally, we emphasize that Theorem 1.11 and Theorem 1.16 make a contribution to the famous Furstenberg–Bergelson–Leibman conjecture, which we now discuss.

1.3. Contributions to the Furstenberg–Bergelson–Leibman conjecture

Furstenberg's ergodic proof [27] of Szemerédi's theorem [59] (on the existence arbitrarily long arithmetic progressions in subsets of integers with positive density) was a departure point for modern ergodic Ramsey theory. We refer to the survey articles [7], [8] and [25], where details (including comprehensive historical background) and an extensive literature are given about this fascinating subject. Ergodic Ramsey theory is a very rich body of research, consisting of many natural generalizations of Szemerédi's theorem, including the celebrated polynomial Szemerédi theorem of Bergelson and Leibman [9] that motivates the following far-reaching conjecture:

Conjecture 1.20 (Furstenberg–Bergelson–Leibman conjecture [10, Section 5.5, p. 468]). *For given parameters $d, k, n \in \mathbb{N}$, let $T_1, \ldots, T_d : X \to X$ be a family of invertible measure-preserving transformations of a probability measure space $(X, \mathcal{B}(X), \mu)$ that generates a nilpotent group of step $l \in \mathbb{Z}_+$, and assume that $P_{1,1}, \ldots, P_{i,j}, \ldots, P_{d,n} \in \mathbb{Z}[m_1, \ldots, m_k]$. Then for any $f_1, \ldots, f_n \in L^\infty(X)$, the nonconventional multiple polynomial averages*

$$A^{P_{1,1},\ldots,P_{d,n}}_{M;X,T_1,\ldots,T_d}(f_1,\ldots,f_n)(x) = \mathbb{E}_{m \in \prod_{j=1}^k [M_j]} \prod_{j=1}^n f_j(T_1^{P_{1,j}(m)} \cdots T_d^{P_{d,j}(m)} x) \tag{1.21}$$

converge for μ-almost every $x \in X$ as $\min\{M_1, \ldots, M_k\} \to \infty$.

Variants of this conjecture were promoted in person by Furstenberg (we refer to Austin's article [3, pp. 6662]) before it was published by Bergelson and Leibman [10, Section 5.5, pp. 468] for $k = 1$. The nilpotent and multi-parameter setting is the appropriate setting for Conjecture 1.20 as convergence may fail if the transformations $T_1, \ldots, T_d$ generate a solvable group, as shown by Bergelson and Leibman [10]. The $L^2(X)$ norm convergence of (1.21) has been studied since Furstenberg's ergodic proof [27] of Szemerédi's theorem [59] and is fairly well-understood (even in the setting of nilpotent groups) due to the groundbreaking work of Walsh [70] with $M_1 = \ldots = M_k$. Prior to Walsh's paper, extensive efforts had been made towards understanding $L^2(X)$ norm convergence, including breakthrough works of Host–Kra [31], Ziegler [71], Bergelson [6] and Leibman [43]. For more details and references, we also refer to [4, 21, 26, 32, 61] and the survey articles [7, 8, 25].

The situation is dramatically different for the pointwise convergence problem (1.21), but recently, significant progress has been made towards establishing the Furstenberg–Bergelson–Leibman conjecture. Now let us make a few remarks about this conjecture, its history and the current state of the art.

1. The case $d = k = n = 1$ of Conjecture 1.20 with $P_{1,1}(m) = m$ follows from Birkhoff's ergodic theorem [11]. In fact, the almost everywhere limit (as well as the norm limit; see also [64]) of (1.21) exists also for all functions $f \in L^p(X)$, with $1 \le p < \infty$, defined on any σ-finite measure space $(X, \mathcal{B}(X), \mu)$.

2. The case $d = k = n = 1$ of Conjecture 1.20 with arbitrary polynomials $P_{1,1} \in \mathbb{Z}[m]$ (as we have seen above) was the famous open problem of Bellow [5] and Furstenberg [24], which was solved by the first author [13, 14, 15] in the mid 1980s. In fact, in [13, 14, 15], it was shown that the almost everywhere limit (as well as the norm limit; see also [29]) of (1.21) exists also for all functions $f \in L^p(X)$, with $1 < p < \infty$, defined on any σ-finite measure space $(X, \mathcal{B}(X), \mu)$. In contrast to the

Birkhoff theorem, if $P_{1,1} \in \mathbb{Z}[n]$ 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 [18] for $P_{1,1}(m) = m^2$ and by LaVictoire [42] for $P_{1,1}(m) = m^k$ for any $k \geq 2$.

3. In the commutative case (step $\ell = 1$) where $d, k \in \mathbb{Z}_+$ and $n = 1$ of Conjecture 1.20 with arbitrary polynomials $P_{1,1}, \ldots, P_{d,1} \in \mathbb{Z}[m_1, \ldots, m_k]$ in the diagonal setting $M_1 = \ldots = M_k$ – that is, the multi-dimensional one-parameter setting – was solved by the second author with Trojan in [54]. As before, it was shown that the almost everywhere limit (as well as the norm limit) of (1.21) exists also for all functions $f \in L^p(X)$, with $1 < p < \infty$, defined on any σ-finite measure space $(X, \mathcal{B}(X), \mu)$.

4. The question to what extent one can relax the commutation relations between $T_1, \ldots, T_d$ in (1.21), even in the one-parameter case $M_1 = \ldots = M_k$, is very intriguing. Some particular examples of averages (1.21) with $d, k \in \mathbb{Z}_+$ and $n = 1$ and polynomial mappings with degree at most two in the step two nilpotent setting were studied in [33, 45]. Recently, the second author with Ionescu, Magyar and Szarek [36] established Conjecture 1.20 with $d \in \mathbb{Z}_+$ and $k = n = 1$ and arbitrary polynomials $P_{1,1}, \ldots, P_{d,1} \in \mathbb{Z}[m]$ in the nilpotent setting (i.e., when $T_1, \ldots, T_d : X \rightarrow X$ 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).

5. In contrast to the commutative linear theory, the multilinear theory is wide open. Only a few results are known in the bilinear $n = 2$ and commutative $d = k = 1$ setting. The first author [16] established pointwise convergence when $P_{1,1}(m) = am$ and $P_{1,2}(m) = bm$, with $a, b \in \mathbb{Z}$. Recently, the third author with Krause and Tao [41] proved pointwise convergence for the polynomial Furstenberg–Weiss averages [28, 30] corresponding to $P_{1,1}(m) = m$ and $P_{1,2}(m) = P(m)$ with $P \in \mathbb{Z}[m]$ and $\deg P \geq 2$.

6. A genuinely multi-parameter case $d = k \geq 2$ with $n = 1$ of Conjecture 1.20 for averages (1.21) with linear orbits (i.e. $P_{j,1}(m_1, \ldots, m_d) = m_j$ for $j \in [d]$) was established independently by Dunford [23] and Zygmund [72] in the early 1950s. Moreover, it follows from [23, 72] that the almost everywhere convergence (as well as the norm convergence) of (1.21) holds for all functions $f \in L^p(X)$, with $1 < p < \infty$, defined on any σ-finite measure space $(X, \mathcal{B}(X), \mu)$ equipped with a family of measure-preserving transformations $T_1, \ldots, T_d : X \rightarrow X$, which does not need to be commutative. One also knows that pointwise convergence fails if $p = 1$. A polynomial variant of the Dunford and Zygmund theorem was discussed above; see (1.7).

We close this discussion by emphasizing that Theorem 1.11 and Theorem 1.16 also contribute to the Furstenberg–Bergelson–Leibman conjecture and, together with all the results listed above, support the evidence that Conjecture 1.20 may be true in full generality though a complete solution seems very difficult.

1.4. Overview of the paper

The paper is organized as follows. In Section 2, we fix necessary notation and terminology. We also introduce the definition of multi-parameter oscillations (2.9) and collect their useful properties; see Proposition 2.16 and Proposition 2.18. In Section 3, we give a detailed proof of Theorem 1.11 by reducing the matter to oscillation estimates for truncated variants of averages $A^P_{M_1, M_2; X} f$; see definition (3.6) and Theorem 3.16, which in turn is reduced to the integer shift system, and Theorem 3.19. A result that may be of independent interest is Proposition 3.7, which shows that oscillations for $A^P_{M_1, M_2; X} f$ and their truncated variants are, in fact, comparable. In Section 3, see inequality (3.5), and we also illustrate how to prove Theorem 1.11 in the degenerate case in the sense of definition (1.14) stated after Theorem 1.11. These arguments can be also used to prove oscillation bounds for the polynomial Dunford and Zygmund theorem, which in turn imply (1.7).

We start with a brief overview of the proof of Theorem 3.19, which implies Theorem 1.5 when $k = 2$ and takes up the bulk of this paper. The proof requires substantial new ideas to overcome a series of new difficulties arising in the multi-parameter setting. These complications do not arise in the

one-parameter setup [13, 14, 15]. The most notable obstacle is the lack of nestedness in the definition of averaging operators (1.8) when the parameters $M_1, \ldots, M_k$ are allowed to run independently. The lack of nestedness complicates every argument in the circle method, which is the main tool in these kinds of problems. In order to understand how the lack of nestedness may affect the underlying arguments, it will be convenient to illustrate this phenomenon by comparing Figure 1 and Figure 2 below. The first picture (Figure 1) represents the family of nested cubes, which is increasing when the time parameter increases. The diagonal relation between parameters $M_1 = \ldots = M_k$ is critical.

The second picture (Figure 2) represents the family which is genuinely multi-parameter and there is no nestedness as the parameters $M_1, \ldots, M_k$ vary independently.

Our remedy to overcome the lack of nestedness will be to develop the so-called multi-parameter circle method, which will be based on an iterative implementation of the classical circle method. Although this idea sounds very simple, it is fairly challenging to formalize it in the context of Conjecture 1.3. We remark that the multi-parameter circle method has been developed for many years in the context of various problems arising in number theory (see [1] for more details and references, including a comprehensive historical background) though it is not applicable directly in the ergodic context. We now highlight the key ingredients that we develop in this paper and that will lead us to develop the multi-parameter circle method in the context of Theorem 3.19:

(i) 'Backwards' Newton diagram is the key tool allowing us to overcome the problem with the lack of nestedness. In particular, it permits us to understand geometric properties of the underlying polynomials in Theorem 3.19 by extracting dominating monomials. The latter are critical in making a distinction between minor and major arcs in the multi-parameter circle method. As far as we know, this is the first time when the concept of Newton diagrams is exploited in problems concerning pointwise ergodic theory. We refer to Section 4 for details.

(ii) We derive new estimates for multi-parameter exponential sums arising in the analysis of Fourier multipliers corresponding to averages (1.10). In Section 5, we build a theory of double exponential sums, which is dictated by the geometry of the corresponding 'backwards' Newton diagrams. Although the theory of multi-parameter exponential sums is rich (see for example, [1]), our results seem to be new and the idea of exploiting 'backwards' Newton diagrams and iterative applications of the Vinogradov mean value theorem [17] in estimates of exponential sums is quite efficient.

(iii) A multi-parameter Ionescu–Wainger multiplier theory is developed in Section 6. The Ionescu–Wainger multiplier theorem [34] was originally proved for linear operators; see also [46, 52, 55, 62]. In this paper, we prove a semi-norm variant of the Ionescu–Wainger theory in the one-parameter setting, which is consequently upgraded to the multi-parameter setup. 'Backwards' Newton diagrams play an essential role in our considerations here as well.

(iv) Finally, we arrive at the stage where the multi-parameter circle method is feasible by a delicate iterative application of the classical circle method. In this part of the argument, the lack of nestedness is particularly unpleasant, causing serious difficulties in controlling error terms that arise in estimating contributions of the corresponding Fourier multipliers on minor and major arcs, which are genuinely multi-parameter. In Section 7, we illustrate how one can use all the tools developed in the previous sections to give a rigorous proof of Theorem 3.19.

We now take a closer look at the tools highlighted above. In Section 4, we introduce the concept of 'backwards' Newton diagram, which is the key to circumvent the difficulties caused by the lack of nestedness. The 'backwards' Newton diagram splits the parameter space into a finite number of sectors, where certain relations between parameters are given. In each of these sectors there is a dominating monomial which, in turn, gives rise to an implementation of the circle method to each of the sectors separately. The distinctions between minor and major arcs are then dictated by the degree of the associated dominating monomial. At this stage, we eliminate minor arcs by invoking estimates of double exponential sums from Proposition 5.37. This proposition is essential in our argument; its proof is given in Section 5. The key ingredients are Proposition 5.22, which may be thought of as a two parameter counterpart of the classical Weyl's inequality, and the properties of the 'backwards' Newton

diagram. Although the theory of multi-parameter exponential sums has been developed over the years (see [1] for a comprehensive treatment of the subject), we require more delicate estimates than those available in the existing literature. In this paper, we give an ad-hoc proof of Proposition 5.22, which follows from an iterative application of Vinogradov's mean value theorem, and may be interesting in its own right. In Section 5, we also develop estimates for complete exponential sums. In Section 6, we develop the Ionescu–Wainger multiplier theory for various semi-norms in one-parameter as well as in multi-parameter settings. Our result in the one-parameter setting, Theorem 6.14, is formulated for oscillations and maximal functions, but the proofs also work for ρ-variations or jumps. In fact, Theorem 6.14 is the starting point for establishing the corresponding multi-parameter Ionescu–Wainger theory for oscillations. The latter theorem will be directly applicable in the analysis of multipliers associated with the averages $A_{M_1,M_2;X}^P f$. The results of Section 6 are critical in our multi-parameter circle method that is presented in Section 7, as it allows us to efficiently control the error terms that arise on major arcs as well as the contribution coming from the main part. In contrast to the one-parameter theory [13, 14, 15], the challenge here is to control, for instance, maximal functions corresponding to error terms. For this purpose, all error terms have to be provided with asymptotic precision, which usually requires careful arguments. The details of the multi-parameter circle method are presented in Section 7 in the context of the proof of Theorem 3.19.

1.5. More about Conjecture 1.20

Conjecture 1.20 is one of the major open problems in pointwise ergodic theory, which seems to be very difficult due to its multilinear nature. Here, in light of the Arkhipov, Chubarikov and Karatsuba [2] equidistribution theory which works also for multiple polynomials, it seems reasonable to propose a slightly more modest problem (implied by Conjecture 1.20) though still very interesting and challenging that can be subsumed under the following conjecture:

Conjecture 1.22. *Let $d, k \in \mathbb{Z}_+$ be given and let $(X, \mathcal{B}(X), \mu)$ be a probability measure space endowed with a family of invertible commuting measure-preserving transformations $T_1, \ldots, T_d : X \to X$. Assume that $P_1, \ldots, P_d \in \mathbb{Z}[m_1, \ldots, m_k]$. Then for any $f \in L^\infty(X)$, the multi-parameter linear polynomial averages*

$$A_{M_1,\ldots,M_k;X,T_1,\ldots,T_d}^{P_1,\ldots,P_d} f(x) = \mathbb{E}_{m \in \prod_{j=1}^k [M_j]} f(T_1^{P_1(m)} \cdots T_d^{P_d(m)} x)$$

converge for μ-almost every $x \in X$, as $\min\{M_1, \ldots, M_k\} \to \infty$.

Even though we prove Conjecture 1.3 here, it is not clear whether Conjecture 1.22 is true for all polynomials. If it is not true for all polynomials, it would be interesting, in view of Theorem 1.16, to characterize the class of those polynomials for which Conjecture 1.22 holds. Although the averages from Theorem 1.11 and Theorem 1.16 share a lot of difficulties that arise in the general case, there are some cases that are not covered by the methods of this paper. An interesting difficulty arises for the so-called partially complete exponential sums when we are seeking estimates of the form

$$\frac{1}{M_1 q} \sum_{m_1=1}^{M_1} \left| \sum_{m_2=1}^q e(a_2 m_2/q + a_3 P(m_1, m_2)/q) \right| \lesssim q^{-\delta}, \tag{1.23}$$

for all $M_1, q \in \mathbb{Z}_+$ and some $\delta \in (0, 1)$, whenever $(a_2, a_3, q) = 1$. These kinds of estimates arise from applications of the circle method with respect to the second variable m_2 for the averages $A_{M_1,M_2;X}^{m_1,m_2,P(m_1,m_2)} f$ when we are at the stage of applying the circle method with respect to the first variable m_1. Here, the assumption that P has partial degrees (as a polynomial of the variable m_1 and a polynomial of the variable m_2) at least two is essential. Otherwise, if $M_1 < q$, the decay $q^{-\delta}$ in (1.23) is not possible. In order to see this, it suffices to take $P(m_1, m_2) = m_1^2 m_2$. A proof of Theorem 1.16 for polynomials

of this type, as well as Conjecture 1.22, will require a deeper understanding and substantially new methods. We believe that the proof of Theorem 1.11 is an important contribution towards understanding Conjecture 1.22 that may shed new light on the general case and either lead to its full resolution or to a counterexample. The second and fourth authors plan to pursue this problem in the future.

1.6. In Memoriam

It was a great privilege and an unforgettable experience for the second and fourth authors to know and work with Elias M. Stein (January 13, 1931–December 23, 2018) and Jean Bourgain (February 28, 1954–December 22, 2018). Eli and Jean had an immeasurable effect on our lives and careers. It was a very sad time for us when we learned that Eli and Jean passed away within an interval of one day in December 2018. We miss our friends and collaborators dearly.

We now briefly describe how the collaboration on this project arose. In 2011, the second and fourth authors started to work on some aspects of a multi-parameter circle method in the context of various discrete multi-parameter operators. These efforts resulted in a draft on estimates for certain two-parameter exponential sums. This draft was sent to the first author sometime in the first part of 2016. In October 2016, when the second author was a member of the Institute for Advanced Study, it was realized (during a discussion between the first two authors) that the estimates from this draft are closely related to a multi-parameter Vinogradov's mean value theorem. This was interesting to the first author who at that time was involved in developing the theory of decoupling. We also realized that some ideas of a multi-parameter circle method from the draft of the second and fourth authors may be upgraded and used in attacking a multi-parameter variant of the Bellow and Furstenberg problem formulated in Conjecture 1.3. That was the first time when the second, third and fourth authors learned about this conjecture and unpublished observations of the first author from the late 1980s that resulted in establishing pointwise convergence in (1.7). This was the starting point of our collaboration. At that time another question arose, which is also related to this paper. It is interesting whether a sharp multi-parameter variant of Vinogradov's mean value theorem can be proved using the recent developments in the decoupling theory from [17]. A multi-parameter Vinogradov's mean value theorem was investigated in [1], but the bounds are not optimal. So the question is about adapting the methods from [17] to the multi-parameter setting in order to obtain sharp bounds, and their applications in the exponential sum estimates.

A substantial part of this project was completed at the end of November/beginning of December 2016, when the fourth author visited Princeton University and the Institute for Advanced Study. At that time, we discussed (more or less) all tools that were needed to establish Theorem 1.11 for the monomial $P(m_1, m_2) = m_1^2 m_2^3$. Then we were convinced that we could establish Conjecture 1.22 in full generality, but various difficulties arose when we started to work out the details, and we ultimately only managed to prove Theorem 1.11 and Theorem 1.16. The second and fourth authors decided to illustrate the arguments in the two-parameter setting and the reason is twofold. On the one hand, we wanted to avoid introducing heavy multi-parameter notation capturing all combinatorial nuances arising in this project. On the other hand, what is more important is that we wanted to illustrate the spirit of our discussions that took place in 2016. For instance, the arguments presented in Section 5 can be derived by using Weyl differencing argument, which may be even simpler and can be easily adapted to the multi-parameter setting, though our presentation is very close to the arguments that we developed in 2016, and also motivates the question about the role of decoupling theory in the multi-parameter Vinogradov's mean value theorem that we have stated above.

2. Notation and useful tools

We now set up notation that will be used throughout the paper. We also collect useful tools and basic properties of oscillation semi-norms that will be used in the paper.

2.1. Basic notation

The set of positive integers and nonnegative integers will be denoted, respectively, by $\mathbb{Z}_+ := \{1, 2, \ldots\}$ and $\mathbb{N} := \{0, 1, 2, \ldots\}$. 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. For any $x \in \mathbb{R}$, we will use the floor and fractional part functions

$$\lfloor x \rfloor := \max\{n \in \mathbb{Z} : n \le x\}, \qquad \text{and} \qquad \{x\} := x - \lfloor x \rfloor.$$

For $x, y \in \mathbb{R}$, we shall also write $x \vee y := \max\{x, y\}$ and $x \wedge y := \min\{x, y\}$. We denote $\mathbb{R}_+ := (0, \infty)$, and for every $N \in \mathbb{R}_+$, we set

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

and we will also write

$$\mathbb{N}_{\le N} := [0, N] \cap \mathbb{N}, \qquad \text{and} \qquad \mathbb{N}_{<N} := [0, N) \cap \mathbb{N},$$
$$\mathbb{N}_{\ge N} := [N, \infty) \cap \mathbb{N}, \qquad \text{and} \qquad \mathbb{N}_{>N} := (N, \infty) \cap \mathbb{N}.$$

For any $\tau > 1$, we will consider the set

$$\mathbb{D}_\tau := \{\tau^n : n \in \mathbb{N}\}.$$

For $a = (a_1, \ldots, a_n) \in \mathbb{Z}^n$ and $q \ge 1$ an integer, we denote by (a, q) the greatest common divisor of a and q; that is, the largest integer $d \ge 1$ that divides q and all the components $a_1, \ldots, a_n$. Clearly, any vector in $\mathbb{Q}^n$ has a unique representation as a/q with $q \in \mathbb{Z}_+$, $a \in \mathbb{Z}^n$ and $(a, q) = 1$.

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, $\mathbb{1}_A(x) = \mathbb{1}_{x \in A}$.

Throughout the paper, $C > 0$ is an absolute constant which may change from occurrence to occurrence. For two nonnegative quantities A, B, we write $A \lesssim B$ if there is an absolute constant $C > 0$ such that $A \le CB$. 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 a function $f : X \to \mathbb{C}$ and positive-valued function $g : X \to (0, \infty)$, we write $f = O(g)$ if there exists a constant $C > 0$ such that $|f(x)| \le Cg(x)$ for all $x \in X$. We will also write $f = O_\delta(g)$ if the implicit constant depends on δ.

2.2. Summation by parts

For any real numbers $u < v$ and any sequences $(a_n : n \in \mathbb{Z}) \subseteq \mathbb{C}$ and $(b_n : n \in \mathbb{Z}) \subseteq \mathbb{C}$, we will use the following version of the summation by parts formula

$$\sum_{n \in (u, v] \cap \mathbb{Z}} a_n b_n = S_v b_{\lfloor v \rfloor} + \sum_{n \in (u, v-1] \cap \mathbb{Z}} S_n (b_n - b_{n+1}), \tag{2.1}$$

where $S_w := \sum_{k \in (u, w] \cap \mathbb{Z}} a_k$ for any $w > u$.

2.3. Euclidean spaces

For $d \in \mathbb{Z}_+$, the set $\{e_i \in \mathbb{R}^d : i \in [d]\}$ denotes the standard basis in $\mathbb{R}^d$. The standard inner product and the corresponding Euclidean norm on $\mathbb{R}^d$ are denoted by

$$x \cdot \xi := \sum_{k=1}^{d} x_k \xi_k, \qquad \text{and} \qquad |x| := |x|_2 := \sqrt{x \cdot x}$$

for every $x = (x_1, \ldots, x_d)$ and $\xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d$.

Throughout the paper, the d-dimensional torus $\mathbb{T}^d$, which unless otherwise stated will be identified with $[-1/2, 1/2)^d$, is a priori endowed with the periodic norm

$$\|\xi\| := \Big(\sum_{k=1}^d \|\xi_k\|^2 \Big)^{1/2} \qquad \text{for} \qquad \xi = (\xi_1, \dots, \xi_d) \in \mathbb{T}^d,$$

where $\|\xi_k\| = \mathrm{disr}(\xi_k, \mathbb{Z})$ for all $\xi_k \in \mathbb{T}$ and $k \in [d]$. However, identifying $\mathbb{T}^d$ with $[-1/2, 1/2)^d$, we see that the norm $\|\cdot\|$ coincides with the Euclidean norm $|\cdot|$ restricted to $[-1/2, 1/2)^d$.

2.4. Smooth functions

The partial derivative of a differentiable function $f : \mathbb{R}^d \to \mathbb{C}$ with respect to the j-th variable x_j will be denoted by $\partial_{x_j} f = \partial_j f$, while for any multi-index $\alpha \in \mathbb{N}^d$, let $\partial^\alpha f$ denote the derivative operator $\partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d} f = \partial_1^{\alpha_1} \cdots \partial_d^{\alpha_d} f$ of total order $|\alpha| := \alpha_1 + \dots + \alpha_d$.

Let $\eta : \mathbb{R} \to [0, 1]$ be a smooth and even cutoff function such that

$$\mathbb{1}_{[-1,1]} \leq \eta \leq \mathbb{1}_{[-2,2]}.$$

For any $n, \xi \in \mathbb{R}$, we define

$$\eta_{\leq n}(\xi) := \eta(2^{-n}\xi).$$

For any $\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d$ and $i \in [d]$, we also define

$$\eta_{\leq n}^{(i)}(\xi) := \eta_{\leq n}(\xi_i).$$

More generally, for any $A = \{i_1, \dots, i_m\} \subseteq [d]$ for some $m \in [d]$, and numbers $n_{i_1}, \dots, n_{i_m} \in \mathbb{R}$ corresponding to the set A, we will write

$$\eta_{\leq n_{i_1}, \dots, \leq n_{i_m}}^A (\xi) := \prod_{j=1}^m \eta_{\leq n_{i_j}}(\xi_{i_j}) = \prod_{j=1}^m \eta_{\leq n_{i_j}}^{(i_j)}(\xi). \tag{2.2}$$

If the elements of the set A are ordered increasingly $1 \leq i_1 < \dots < i_m \leq d$, we will also write

$$\eta_{\leq n_{i_1}, \dots, \leq n_{i_m}}^{(i_1, \dots, i_m)}(\xi) := \eta_{\leq n_{i_1}, \dots, \leq n_{i_m}}^A(\xi) = \prod_{j=1}^m \eta_{\leq n_{i_j}}(\xi_{i_j}) = \prod_{j=1}^m \eta_{\leq n_{i_j}}^{(i_j)}(\xi).$$

If $n_{i_1} = \dots = n_{i_m} = n \in \mathbb{R}$, we will abbreviate $\eta_{\leq n_{i_1}, \dots, \leq n_{i_m}}^A$ to $\eta_{\leq n}^A$ and $\eta_{\leq n_{i_1}, \dots, \leq n_{i_m}}^{(i_1, \dots, i_m)}$ to $\eta_{\leq n}^{(i_1, \dots, i_m)}$.

2.5. Function spaces

All vector spaces in this paper will be defined over the complex numbers $\mathbb{C}$. The triple $(X, \mathcal{B}(X), \mu)$ is a measure space X with σ-algebra $\mathcal{B}(X)$ and σ-finite measure μ. The space of all μ-measurable complex-valued functions defined on X 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 finite dimensional normed vector space $(B, \|\cdot\|_B)$; for instance,

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

where $L^0(X; B)$ denotes the space of measurable functions from X to B (up to almost everywhere equivalence). Of course, if B is separable, these notions can be extended to infinite-dimensional B. In this paper, we will always be able to work in finite-dimensional settings by appealing to standard approximation arguments. In our case, we will usually have $X = \mathbb{R}^d$ or $X = \mathbb{T}^d$ equipped with Lebesgue measure, and $X = \mathbb{Z}^d$ endowed with counting measure. If X is endowed with counting measure, we will abbreviate $L^p(X)$ to $\ell^p(X)$ and $L^p(X; B)$ to $\ell^p(X; B)$.

If $T : B_1 \to B_2$ is a continuous linear map between two normed vector spaces B_1 and B_2, we use $\|T\|_{B_1 \to B_2}$ to denote its operator norm.

The following extension of the Marcinkiewicz–Zygmund inequality to the Hilbert space setting will be very useful in Section 6.

Lemma 2.3. *Let $(X, \mathcal{B}(X), \mu)$ be a σ-finite measure space endowed with a family $T = (T_m : m \in \mathbb{N})$ of bounded linear operators $T_m : L^p(X) \to L^p(X)$ for some $p \in (0, \infty)$. Suppose that*

$$A_p(T) := \sup_{(\omega_m : m \in \mathbb{N}) \in \{-1, 1\}^{\mathbb{N}}} \Big\| \sum_{m \in \mathbb{N}} \omega_m T_m \Big\|_{L^p \to L^p} < \infty.$$

Then there is a constant $C_p > 0$ such that for every sequence $(f_j : j \in \mathbb{N}) \in L^p(X; \ell^2(\mathbb{N}))$, we have

$$\Big\| \Big(\sum_{j \in \mathbb{N}} \sum_{m \in \mathbb{N}} |T_m f_j|^2 \Big)^{1/2} \Big\|_{L^p(X)} \le C_p A_p(T) \Big\| \Big(\sum_{j \in \mathbb{N}} |f_j|^2 \Big)^{1/2} \Big\|_{L^p(X)}. \tag{2.4}$$

The index set $\mathbb{N}$ in the inner sum of (2.4) can be replaced by any other countable set and the result remains valid.

The proof of Lemma 2.3 can be found in [48].

2.6. Fourier transform

We shall write $e(z) = e^{2\pi i z}$ 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

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

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

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

Sometimes we shall abbreviate $\mathcal{F}_{\mathbb{Z}^d} f$ to $\hat{f}$.

Let $\mathbb{G} = \mathbb{R}^d$ or $\mathbb{G} = \mathbb{Z}^d$. The 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 $\mathfrak{m} : \mathbb{G}^* \to \mathbb{C}$ and a test function $f : \mathbb{G} \to \mathbb{C}$, we define the Fourier multiplier operator by

$$T_{\mathbb{G}}[\mathfrak{m}] f(x) := \int_{\mathbb{G}^*} e(-\xi \cdot x) \mathfrak{m}(\xi) \mathcal{F}_{\mathbb{G}} f(\xi) d\xi, \quad \text{for} \quad x \in \mathbb{G}. \tag{2.5}$$

One may think that $f : \mathbb{G} \to \mathbb{C}$ is a compactly supported function on $\mathbb{G}$ (and smooth if $\mathbb{G} = \mathbb{R}^d$) or any other function for which (2.5) makes sense.

Let $\mathbb{R}_{\leq d}[x_1, \ldots, x_n]$ be the vector space of all polynomials on $\mathbb{R}^n$ of degree at most $d \in \mathbb{Z}_+$, which is equipped with the norm $\|P\| := \sum_{0 \leq |\beta| \leq d} |c_\beta|$ whenever

$$P(x) = \sum_{0 \leq |\beta| \leq d} c_\beta x_1^{\beta_1} \cdots x_n^{\beta_n} \quad \text{for} \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

We now formulate a multidimensional variant of the van der Corput lemma for polynomials that will be useful in our further applications.

Proposition 2.6. *For each $d, n \in \mathbb{Z}_+$, there exists a constant $C_{d,n} > 0$ such that for any $P \in \mathbb{R}_{\leq d}[x_1, \ldots, x_n]$ with $P(0) = 0$, one has*

$$\left| \int_{[0,1]^n} e(P(x)) dx \right| \leq C_{d,n} \|P\|^{-1/d}.$$

The proof of Proposition 2.6 can be found in [20, Corollary 7.3., p. 1008]; see also [1, Section 1].

2.7. Comparing sums to integrals

A well-known but useful lemma comparing sums to integrals is the following. The proof can be found in [73, Chapter V]; see also [63].

Lemma 2.7. *Suppose $f : [a, b] \to \mathbb{R}$ is C^1 such that f' is monotonic and $|f'(s)| \leq 1/2$ on $[a, b]$. Then there is an absolute constant A such that*

$$\left| \sum_{a < n \leq b} e(f(n)) - \int_a^b e(f(s)) ds \right| \leq A.$$

2.8. Coordinatewise order $\preceq$

For any $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$ and $y = (y_1, \ldots, y_k) \in \mathbb{R}^k$, we say $x \preceq y$ if and only if $x_i \leq y_i$ for each $i \in [k]$. We also write $x \prec y$ if and only if $x \preceq y$ and $x \neq y$, and $x \prec_s y$ 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 $\#\mathbb{I} \geq 2$ and for every $J \in \mathbb{Z}_+ \cup \{\infty\}$, define the set

$$\mathfrak{S}_J(\mathbb{I}) := \left\{ (t_i : i \in \mathbb{N}_{\leq J}) \subseteq \mathbb{I} : t_0 \prec_s t_1 \prec_s \ldots \prec_s t_J \right\}, \tag{2.8}$$

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

2.9. Oscillation semi-norms

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

$$O_{I,J}(\mathfrak{a}_t(x) : t \in \mathbb{J}) := \left(\sum_{j=0}^{J-1} \sup_{t \in \mathbb{B}[I,j] \cap \mathbb{J}} \left| \mathfrak{a}_t(x) - \mathfrak{a}_{I_j}(x) \right|^2 \right)^{1/2}, \tag{2.9}$$

where $\mathbb{B}[I, i] := [I_{i1}, I_{(i+1)1}) \times \ldots \times [I_{ik}, I_{(i+1)k})$ is a box determined by the element $I_i = (I_{i1}, \ldots, I_{ik})$ 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 2.10. Some remarks concerning the definition of oscillation semi-norms are in order.

1. Clearly, $O_{I,J}(\mathfrak{a}_t : t \in \mathbb{J})$ 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

$$O_{I,J}(\mathfrak{a}_t : t \in \mathbb{J}_1 \cup \mathbb{J}_2) \leq O_{I,J}(\mathfrak{a}_t : t \in \mathbb{J}_1) + O_{I,J}(\mathfrak{a}_t : t \in \mathbb{J}_2). \tag{2.11}$$

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

$$O_{I,J}(\mathfrak{a}_t : t \in \mathbb{J}) \lesssim \Big(\sum_{t \in \mathbb{I}} |\mathfrak{a}_t|^2 \Big)^{1/2}.$$

4. Let $\mathbb{I} \subseteq \mathbb{R}^k$ be a countable index set such that $\#\mathbb{I} \geq 2$. For $l \in [k]$, let $p_l : \mathbb{R}^k \to \mathbb{R}$ be the lth coordinate projection. Note that for any family $(\mathfrak{a}_t : t \in \mathbb{I}) \subseteq \mathbb{C}$, any $J \in \mathbb{Z}_+$, any $I \in \mathfrak{S}_J(\mathbb{I})$ and any $l \in [k]$, one has

$$O_{I,J}(\mathfrak{a}_t : t \in \mathbb{I}) = \Big(\sum_{j=0}^{J-1} \sup_{t \in \mathbb{B}[I,j] \cap \mathbb{I}} |\mathfrak{a}_t - \mathfrak{a}_{I_j}|^2 \Big)^{1/2}$$

$$\lesssim \Big(\sum_{t_l \in p_l(\mathbb{I})} \sup_{\substack{(t_1,\ldots,t_{l-1},t_{l+1},\ldots,t_k) \in \prod_{i \in [k] \setminus \{l\}} p_i(\mathbb{I}) \\ (t_1,\ldots,t_{l-1},t_l,t_{l+1},\ldots,t_k) \in \mathbb{I}}} |\mathfrak{a}_{(t_1,\ldots,t_{l-1},t_l,t_{l+1},\ldots,t_k)}|^2 \Big)^{1/2}, \tag{2.12}$$

where $p_l(\mathbb{I}) \subset \mathbb{R}$ is the image of $\mathbb{I}$ under p_l. Inequality (2.12) will be repeatedly used in Section 7. It is important to note that the parameter $t \in \mathbb{I}$ in the definition of oscillations and the sequence $I \in \mathfrak{S}_J(\mathbb{I})$ both take values in $\mathbb{I}$.

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

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

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

It is clear that for any $\mathbb{I} \subseteq \mathbb{R}$ such that $\#\mathbb{I} \geq 2$, any $J \in \mathbb{Z}_+ \cup \{\infty\}$ and any sequence $I = (I_i : i \in \mathbb{N}_{\leq J}) \in \mathfrak{S}_J(\mathbb{I})$, one has

$$O_{I,J}(\mathfrak{a}_t : t \in \mathbb{I}) \leq V^\rho(\mathfrak{a}_t : t \in \mathbb{I}), \tag{2.13}$$

whenever $1 \leq \rho \leq 2$.

6. Inequality (2.13) allows us to deduce the Rademacher–Menshov inequality for oscillations, which asserts that for any $j_0, m \in \mathbb{N}$ so that $j_0 < 2^m$ and any sequence of complex numbers $(\mathfrak{a}_k : k \in \mathbb{N})$, any $J \in [2^m]$ and any $I \in \mathfrak{S}_J([j_0, 2^m))$, we have

$$O_{I,J}(\mathfrak{a}_j : j_0 \leq j < 2^m) \leq V^2(\mathfrak{a}_j : j_0 \leq j < 2^m)$$

$$\leq \sqrt{2} \sum_{i=0}^m \Big(\sum_{j=0}^{2^{m-i}-1} \Big| \sum_{\substack{k \in U_j^i \\ U_j^i \subseteq [j_0, 2^m)}} \mathfrak{a}_{k+1} - \mathfrak{a}_k \Big|^2 \Big)^{1/2}, \tag{2.14}$$

where $U_j^i := [j2^i, (j+1)2^i)$ for any $i, j \in \mathbb{Z}$. The latter inequality in (2.14) immediately follows from [51, Lemma 2.5., p. 534]. Inequality (2.14) will be used in Section 6.

7. For any $p \in [1, \infty]$ and for any family $(\mathfrak{a}_t : t \in \mathbb{N}^k) \subseteq \mathbb{C}$ of k-parameter measurable functions on X, one has

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{N}^k)} \left\| O_{I,J}(\mathfrak{a}_t : t \in \mathbb{N}^k) \right\|_{L^p(X)} \leq 2 \left\| \sup_{t \in \mathbb{N}^k} |\mathfrak{a}_t| \right\|_{L^p(X)}$$
$$+ \sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{Z}_+^k)} \left\| O_{I,J}(\mathfrak{a}_t : t \in \mathbb{Z}_+^k) \right\|_{L^p(X)}. \tag{2.15}$$

This easily follows from the definition of the set $\mathfrak{S}_J(\mathbb{N}^k)$; see (2.8).

8. 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 $\bar{I}_i = (I_i, \ldots, I_i) \in \mathbb{I}^k$ for each $i \in \mathbb{N}_{\leq J}$. Then for any $\mathbb{J} \subseteq \mathbb{I}^k$, one has

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

It is not difficult to show that oscillation semi-norms always dominate maximal functions.

Proposition 2.16. *Assume that $k \in \mathbb{Z}_+$ and let $(\mathfrak{a}_t : t \in \mathbb{R}^k) \subseteq \mathbb{C}$ be a k-parameter family of measurable functions on X. Let $\mathbb{I} \subseteq \mathbb{R}$ and $\#\mathbb{I} \geq 2$. Then for every $p \in [1, \infty]$, we have*

$$\left\| \sup_{t \in (\mathbb{I} \setminus \{\sup \mathbb{I}\})^k} |\mathfrak{a}_t| \right\|_{L^p(X)} \leq \sup_{t \in \mathbb{I}^k} \|\mathfrak{a}_t\|_{L^p(X)} + \sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{I})} \left\| O_{\bar{I},J}(\mathfrak{a}_t : t \in \mathbb{I}^k) \right\|_{L^p(X)}, \tag{2.17}$$

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 2.10.

A remarkable feature of the oscillation semi-norms is that they imply pointwise convergence, which is formulated precisely in the following proposition.

Proposition 2.18. *Let $(X, \mathcal{B}(X), \mu)$ be a σ-finite measure space. For $k \in \mathbb{Z}_+$, let $(\mathfrak{a}_t : t \in \mathbb{N}^k) \subseteq \mathbb{C}$ be a k-parameter family of measurable functions on X. Suppose that there is $p \in [1, \infty)$ and a constant $C_p > 0$ such that*

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{N})} \left\| O_{\bar{I},J}(\mathfrak{a}_t : t \in \mathbb{N}^k) \right\|_{L^p(X)} \leq C_p < \infty,$$

where $\bar{I} \in \mathfrak{S}_J(\mathbb{N}^k)$ is the diagonal sequence corresponding to a sequence $I \in \mathfrak{S}_J(\mathbb{N})$ as in Remark 2.10. Then the limit

$$\lim_{\min\{t_1, \ldots, t_k\} \to \infty} \mathfrak{a}_{(t_1, \ldots, t_k)}$$

exists μ-almost everywhere on X.

For detailed proofs of Proposition 2.16 and Proposition 2.18, we refer to [53].

3. Basic reductions and ergodic theorems: Proof of Theorem 1.11

This section is intended to establish Theorem 1.11 for general measure-preserving systems by reducing the matter to the integer shift system. We first briefly explain that the oscillation inequality (1.13) from item (iv) of Theorem 1.11 implies conclusions from items (i)–(iii) of this theorem.

3.1. Proof of Theorem 1.11(iii)

Assuming Theorem 1.11(iv) with $\tau = 2$ and invoking Proposition 2.16 (this permits us to dominate maximal functions by oscillations), we see that for every $p \in (1, \infty)$, there is a constant $C_p > 0$ such

that for any $f \in L^p(X)$, one has

$$\left\| \sup_{M_1,M_2 \in \mathbb{D}_2} |A^P_{M_1,M_2;X} f| \right\|_{L^p(X)} \lesssim_{p,P} \|f\|_{L^p(X)}. \tag{3.1}$$

But for any $f \geq 0$, we have also a simple pointwise bound

$$\sup_{M_1,M_2 \in \mathbb{Z}_+} A^P_{M_1,M_2;X} f \lesssim \sup_{M_1,M_2 \in \mathbb{D}_2} A^P_{M_1,M_2;X} f,$$

which in view of (3.1) gives (1.12) as claimed.

3.2. Proof of Theorem 1.11(ii)

We fix $p \in (1, \infty)$ and $f \in L^p(X)$. We can also assume that $f \geq 0$. Using (1.13) with $\tau = 2^{1/s}$ for every $s \in \mathbb{Z}_+$ and invoking Proposition 2.18, we conclude that there is $f_s^* \in L^p(X)$ such that

$$\lim_{\min\{n_1,n_2\} \to \infty} A^P_{2^{n_1/s},2^{n_2/s};X} f(x) = f_s^*(x)$$

μ-almost everywhere on X for every $s \in \mathbb{Z}_+$. It is not difficult to see that $f_1^* = f_s^*$ for all $s \in \mathbb{Z}_+$, since $\mathbb{D}_2 \subseteq \mathbb{D}_{2^{1/s}}$. Now for each $s \in \mathbb{Z}_+$ and each $M_1, M_2 \in \mathbb{Z}_+$, let $n^i_{M_i} \in \mathbb{N}$ be such that $2^{n^i_{M_i}/s} \leq M_i < 2^{(n^i_{M_i}+1)/s}$ for $i \in [2]$. Then we may conclude

$$2^{-2/s} f_1^*(x) \leq \liminf_{\min\{M_1,M_2\} \to \infty} A^P_{M_1,M_2;X} f(x) \leq \limsup_{\min\{M_1,M_2\} \to \infty} A^P_{M_1,M_2;X} f(x) \leq 2^{2/s} f_1^*(x).$$

Letting $s \to \infty$, we obtain

$$\lim_{\min\{M_1,M_2\} \to \infty} A^P_{M_1,M_2;X} f(x) = f_1^*(x)$$

μ-almost everywhere on X. This completes the proof of Theorem 1.11(ii).

3.3. Proof of Theorem 1.11(i)

Finally, pointwise convergence from Theorem 1.11(ii) combined with the maximal inequality (1.12) and dominated convergence theorem gives norm convergence for any $f \in L^p(X)$ with $1 < p < \infty$. This completes the proof of Theorem 1.11.

3.4. Proof of Theorem 1.11 in the degenerate case

It is perhaps worth remarking that the proof of Theorem 1.11 is fairly easy when $P \in \mathbb{Z}[m_1, m_2]$ is degenerate in the sense that it can be written as $P(m_1, m_2) = P_1(m_1) + P_2(m_2)$, where $P_1 \in \mathbb{Z}[m_1]$ and $P_2 \in \mathbb{Z}[m_2]$ such that $P_1(0) = P_2(0) = 0$ (see (1.14)). It suffices to prove (1.13). The crucial observation is the following identity:

$$A^{P_1(m_1)}_{M_1;X} A^{P_2(m_2)}_{M_2;X} f = A^{P_2(m_2)}_{M_2;X} A^{P_1(m_1)}_{M_1;X} f = A^{P(m_1,m_2)}_{M_1,M_2;X} f. \tag{3.2}$$

Recall from [48] that for every $p \in (1, \infty)$, there is $C_p > 0$ such that for every $f = (f_\iota : \iota \in \mathbb{N}) \in L^p(X; \ell^2(\mathbb{N}))$ and $i \in [2]$, one has

$$\left\| \left(\sum_{\iota \in \mathbb{N}} \sup_{M_i \in \mathbb{Z}_+} |A^{P_i(m_i)}_{M_i;X} f_\iota|^2 \right)^{1/2} \right\|_{L^p(X)} \leq C_p \|f\|_{L^p(X;\ell^2)}. \tag{3.3}$$

Moreover, from [47], it was proved that for every $p \in (1, \infty)$, there is $C_p > 0$ such that for every $f \in L^p(X)$ and $i \in [2]$, one has

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{Z}_+)} \|O_{I,J}(A_{M_i;X}^{P_i(\mathfrak{m}_i)} f : M_i \in \mathbb{Z}_+)\|_{L^p(X)} \leq C_p \|f\|_{L^p(X)}. \tag{3.4}$$

By (3.2), for every $J \in \mathbb{Z}_+$, $I \in \mathfrak{S}_J(\mathbb{Z}_+^2)$ and $j \in \mathbb{N}_{<J}$, one can write

$$\sup_{(M_1,M_2) \in \mathbb{B}[I,j]} \left| A_{M_1,M_2;X}^{P(\mathfrak{m}_1,\mathfrak{m}_2)} f - A_{I_{j1},I_{j2};X}^{P(\mathfrak{m}_1,\mathfrak{m}_2)} f \right| \leq \sup_{M_1 \in \mathbb{Z}_+} \left| A_{M_1;X}^{P_1(\mathfrak{m}_1)} \Big(\sup_{I_{j2} \leq M_2 < I_{(j+1)2}} |A_{M_2;X}^{P_2(\mathfrak{m}_2)} f - A_{I_{j2};X}^{P_2(\mathfrak{m}_2)} f| \Big) \right.$$
$$+ \sup_{M_2 \in \mathbb{Z}_+} \left| A_{M_2;X}^{P_2(\mathfrak{m}_2)} \Big(\sup_{I_{j1} \leq M_1 < I_{(j+1)1}} |A_{M_1;X}^{P_1(\mathfrak{m}_1)} f - A_{I_{j1};X}^{P_1(\mathfrak{m}_1)} f| \Big) \right|.$$

Using this inequality with the vector-valued maximal inequality (3.3) and one-parameter oscillation inequality (3.4), one obtains

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{Z}_+^2)} \left\| O_{I,J}(A_{M_1,M_2;X}^{P(\mathfrak{m}_1,\mathfrak{m}_2)} f : M_1, M_2 \in \mathbb{Z}_+) \right\|_{L^p(X)}$$
$$\lesssim \sum_{i \in [2]} \sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{Z}_+)} \left\| O_{I,J}(A_{M_i;X}^{P_i(\mathfrak{m}_i)} f : M_i \in \mathbb{Z}_+) \right\|_{L^p(X)} \lesssim_p \|f\|_{L^p(X)}. \tag{3.5}$$

This completes the proof of Theorem 1.11 in the degenerate case. From now on we will additionally assume that $P \in \mathbb{Z}[\mathfrak{m}_1, \mathfrak{m}_2]$ is non-degenerate.

3.5. Reductions to truncated averages

We have seen that the proof of Theorem 1.11 has been reduced to proving the oscillation inequality (1.13). We begin with certain general reductions that will simplify our further arguments. Let us fix our measure-preserving transformations $T_1, \ldots, T_d$, our polynomials $\mathcal{P} = \{P_1, \ldots, P_d\} \subset \mathbb{Z}[\mathfrak{m}_1, \ldots, \mathfrak{m}_k]$ and define a truncated version of the average (1.8) by

$$\tilde{A}_{M_1,\ldots,M_k;X}^{\mathcal{P}} f(x) := \mathbb{E}_{m \in R_{M_1,\ldots,M_k}} f(T_1^{P_1(m)} \cdots T_d^{P_d(m)} x), \qquad x \in X, \tag{3.6}$$

where

$$R_{M_1,\ldots,M_k} := ([M_1] \setminus [\tau^{-1} M_1]) \times \cdots \times ([M_k] \setminus [\tau^{-1} M_k])$$

is a rectangle in $\mathbb{Z}^k$.

We will abbreviate $\tilde{A}_{M_1,\ldots,M_k;X}^{\mathcal{P}}$ to $\tilde{A}_{M;X}^{\mathcal{P}}$ and $R_{M_1,\ldots,M_k}$ to R_M whenever $M = (M_1, \ldots, M_k) \in \mathbb{Z}_+^k$. We now show that the $L^p(X)$ norms of the oscillation semi-norms associated with the averages from (1.8) and (3.6) have comparable norms in the following sense.

Proposition 3.7. *Let $d, k \in \mathbb{Z}_+$ be given. Let $(X, \mathcal{B}(X), \mu)$ be a σ-finite measure space equipped with a family of commuting invertible and measure-preserving transformations $T_1, \ldots, T_d : X \to X$. Let $\mathcal{P} = \{P_1, \ldots, P_d\} \subset \mathbb{Z}[\mathfrak{m}_1, \ldots, \mathfrak{m}_k]$, $M = (M_1, \ldots, M_k)$ and let $A_{M;X}^{\mathcal{P}}$ and $\tilde{A}_{M;X}^{\mathcal{P}}$ be the corresponding averaging operators defined respectively in (1.8) and (3.6). For every $\tau > 1$ and every $1 \leq p \leq \infty$, there is a finite constant $C := C_{d,k,p,\tau} > 0$ such that for any $f \in L^p(X)$, one has*

$$\Big\| \sup_{M \in \mathbb{D}_\tau^k} |A_{M;X}^{\mathcal{P}}| \Big\|_{L^p(X)} \leq C \Big\| \sup_{M \in \mathbb{D}_\tau^k} |\tilde{A}_{M;X}^{\mathcal{P}}| \Big\|_{L^p(X)}. \tag{3.8}$$

An oscillation variant of (3.8) also holds

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^k)} \left\| O_{I,J}(A_{M;X}^{\mathcal{P}} : M \in \mathbb{D}_\tau^k) \right\|_{L^P(X)}$$

$$\leq C \sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^k)} \left\| O_{I,J}(\tilde{A}_{M;X}^{\mathcal{P}} f : M \in \mathbb{D}_\tau^k) \right\|_{L^P(X)} + C \|f\|_{L^P(X)}. \tag{3.9}$$

Proof. The proof will proceed in two steps. We begin with some general observations which will permit us to simplify further arguments leading to the proofs of (3.8) and (3.9).

Step 1. Suppose that $(\mathfrak{a}_m : m \in \mathbb{Z}_+^k)$ is a k-parameter sequence of measurable functions on X. Then for $M = (M_1, \ldots, M_k) = (\tau^{n_1}, \ldots, \tau^{n_k}) \in \mathbb{D}_\tau^k$, one can write

$$\sum_{m \in Q_{M_1, \ldots, M_k}} \mathfrak{a}_m = \sum_{(l_1, \ldots, l_k) \in \mathbb{N}_{\leq n_1} \times \cdots \times \mathbb{N}_{\leq n_k}} \sum_{m \in R_{\tau^{l_1}, \ldots, \tau^{l_k}}} \mathfrak{a}_m,$$

and

$$\sum_{(l_1, \ldots, l_k) \in \mathbb{N}_{\leq n_1} \times \cdots \times \mathbb{N}_{\leq n_k}} \frac{|R_{\tau^{l_1}, \ldots, \tau^{l_k}}|}{|Q_{\tau^{n_1}, \ldots, \tau^{n_k}}|} \lesssim_{k, \tau} 1.$$

Combining these two estimates, one sees that

$$\left\| \sup_{M \in \mathbb{D}_\tau^k} |\mathbb{E}_{m \in Q_M} \mathfrak{a}_m| \right\|_{L^P(X)} \lesssim_{k, \tau} \left\| \sup_{M \in \mathbb{D}_\tau^k} |\mathbb{E}_{m \in R_M} \mathfrak{a}_m| \right\|_{L^P(X)}. \tag{3.10}$$

Applying (3.10) with $\mathfrak{a}_m(x) = f(T_1^{P_1(m)} \cdots T_d^{P_d(m)} x)$, we obtain (3.8).

Step 2. As before, let $(\mathfrak{a}_m : m \in \mathbb{Z}_+^k)$ be a k-parameter sequence of measurable functions on X. For $l \in \mathbb{N}_{\leq k}$ and $M = (M_1, \ldots, M_k) = (\tau^{n_1}, \ldots, \tau^{n_k}) \in \mathbb{D}_\tau^k$, define the sets

$$B_M^l := \prod_{i=1}^l ([M_i] \setminus [\tau^{-1} M_i]) \times \prod_{i=l+1}^k [M_i] \quad \text{and} \quad D_M^l := \prod_{i=1}^{l-1} ([M_i] \setminus [\tau^{-1} M_i]) \times [\tau^{-1} M_l] \times \prod_{i=l+1}^k [M_i].$$

Note that $B_M^0 = Q_M$ and $B_M^k = R_M$, and $B_M^{l-1} = B_M^l \cup D_M^l$. Moreover, for $l \in [k]$, one sees

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^k)} \left\| O_{I,J}(\mathbb{E}_{m \in B_M^{l-1}} \mathfrak{a}_m : M \in \mathbb{D}_\tau^k) \right\|_{L^P(X)}$$

$$\leq \sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^k)} \left\| O_{I,J}(u_{M_l} \mathbb{E}_{m \in B_M^l} \mathfrak{a}_m : M \in \mathbb{D}_\tau^k) \right\|_{L^P(X)} \tag{3.11}$$

$$+ \sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^k)} \left\| O_{I,J}(v_{M_l} \mathbb{E}_{m \in D_M^l} \mathfrak{a}_m : M \in \mathbb{D}_\tau^k) \right\|_{L^P(X)},$$

where

$$u_{M_l} := \frac{|B_M^l|}{|B_M^{l-1}|} = \frac{\lfloor M_l \rfloor - \lfloor \tau^{-1} M_l \rfloor}{\lfloor M_l \rfloor} \quad \text{and} \quad v_{M_l} := \frac{|D_M^l|}{|B_M^{l-1}|} = \frac{\lfloor \tau^{-1} M_l \rfloor}{\lfloor M_l \rfloor}.$$

Considering $\tilde{u}_{M_l} := u_{M_l} - 1 + \tau^{-1}$ and $\tilde{v}_{M_l} := v_{M_l} - \tau^{-1}$, we see that

$$\sum_{M_l \in \mathbb{D}_\tau} \tilde{u}_{M_l}^2 \lesssim_\tau 1, \quad \text{and} \quad \sum_{M_l \in \mathbb{D}_\tau} \tilde{v}_{M_l}^2 \lesssim_\tau 1.$$

Thus, using (2.12), one sees that

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^k)} \left\| O_{I,J}(\tilde{u}_{M_l} \mathbb{E}_{m \in B_M^l} \mathfrak{a}_m : M \in \mathbb{D}_\tau^k) \right\|_{L^p(X)} \lesssim_\tau \left\| \sup_{M \in \mathbb{D}_\tau^k} |\mathbb{E}_{m \in Q_M} \mathfrak{a}_m| \right\|_{L^p(X)},$$

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^k)} \left\| O_{I,J}(\tilde{v}_{M_l} \mathbb{E}_{m \in D_M^l} \mathfrak{a}_m : M \in \mathbb{D}_\tau^k) \right\|_{L^p(X)} \lesssim_\tau \left\| \sup_{M \in \mathbb{D}_\tau^k} |\mathbb{E}_{m \in Q_M} \mathfrak{a}_m| \right\|_{L^p(X)}. \tag{3.12}$$

By (2.15), there is $C_{p,\tau} > 0$ such that

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^k)} \left\| O_{I,J}(v_{M_l} \mathbb{E}_{m \in D_M^l} \mathfrak{a}_m : M \in \mathbb{D}_\tau^k) \right\|_{L^p(X)} \le C_{p,\tau} \left\| \sup_{M \in \mathbb{D}_\tau^k} |\mathbb{E}_{m \in Q_M} \mathfrak{a}_m| \right\|_{L^p(X)}$$

$$+ \sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^k)} \left\| O_{I,J}(v_{M_l} \mathbb{E}_{m \in B_M^{l-1}} \mathfrak{a}_m : M \in \mathbb{D}_\tau^k) \right\|_{L^p(X)}. \tag{3.13}$$

Finally, combining (3.11), (3.12) and (3.13), one obtains the following bootstrap inequality:

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^k)} \left\| O_{I,J}(\mathbb{E}_{m \in B_M^{l-1}} \mathfrak{a}_m : M \in \mathbb{D}_\tau^k) \right\|_{L^p(X)} \le C_{p,\tau} \left\| \sup_{M \in \mathbb{D}_\tau^k} |\mathbb{E}_{m \in Q_M} \mathfrak{a}_m| \right\|_{L^p(X)}$$

$$+ \tau^{-1} \sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^k)} \left\| O_{I,J}(\mathbb{E}_{m \in B_M^{l-1}} \mathfrak{a}_m : M \in \mathbb{D}_\tau^k) \right\|_{L^p(X)}$$

$$+ \tau^{-1}(\tau - 1) \sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^k)} \left\| O_{I,J}(\mathbb{E}_{m \in B_M^l} \mathfrak{a}_m : M \in \mathbb{D}_\tau^k) \right\|_{L^p(X)},$$

which immediately yields

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^k)} \left\| O_{I,J}(\mathbb{E}_{m \in B_M^{l-1}} \mathfrak{a}_m : M \in \mathbb{D}_\tau^k) \right\|_{L^p(X)} \le C_{p,\tau} \left\| \sup_{M \in \mathbb{D}_\tau^k} |\mathbb{E}_{m \in Q_M} \mathfrak{a}_m| \right\|_{L^p(X)}$$

$$+ \sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^k)} \left\| O_{I,J}(\mathbb{E}_{m \in B_M^l} \mathfrak{a}_m : M \in \mathbb{D}_\tau^k) \right\|_{L^p(X)}. \tag{3.14}$$

Iterating (3.14) k times and using (3.10) to control the maximal function, we conclude that

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^k)} \left\| O_{I,J}(\mathbb{E}_{m \in Q_M} \mathfrak{a}_m : M \in \mathbb{D}_\tau^k) \right\|_{L^p(X)} \le C_{p,\tau} \left\| \sup_{M \in \mathbb{D}_\tau^k} |\mathbb{E}_{m \in R_M} \mathfrak{a}_m| \right\|_{L^p(X)}$$

$$+ \sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^k)} \left\| O_{I,J}(\mathbb{E}_{m \in R_M} \mathfrak{a}_m : M \in \mathbb{D}_\tau^k) \right\|_{L^p(X)}. \tag{3.15}$$

Finally, using (3.15) with $\mathfrak{a}_m(x) = f(T_1^{P_1(m)} \cdots T_d^{P_d(m)} x)$ and invoking Proposition 2.16 (to control the maximal function from (3.15) by oscillation semi-norms), we obtain (3.9) as desired. □

Now using Proposition 3.7, we can reduce the oscillation inequality (1.13) from Theorem 1.11 to establish the following result for non-degenerate polynomials in the sense of (1.14).

Theorem 3.16. *Let* $(X, \mathcal{B}(X), \mu)$ *be a σ-finite measure space equipped with an invertible measure-preserving transformation* $T : X \to X$. *Let* $P \in \mathbb{Z}[\mathrm{m}_1, \mathrm{m}_2]$ *be a non-degenerate polynomial such that* $P(0, 0) = 0$. *Let* $\tilde{A}_{M;X}^P f$ *with* $M = (M_1, M_2)$ *be the average defined in (3.6) with* $d = 1$, $k = 2$ *and* $P_1 = P$. *If* $1 < p < \infty$ *and* $\tau > 1$, *and* $\mathbb{D}_\tau := \{\tau^n : n \in \mathbb{N}\}$, *then one has*

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^2)} \left\| O_{I,J}(\tilde{A}_{M_1,M_2;X}^P f : M_1, M_2 \in \mathbb{D}_\tau) \right\|_{L^p(X)} \lesssim_{p,\tau,P} \|f\|_{L^p(X)}. \tag{3.17}$$

The implicit constant in (3.17) can be taken to depend only on p, τ, P.

3.6. Reduction to the integer shift system

As mentioned in Example 1.9, the integer shift system is the most important for pointwise convergence problems. For $T = S_1$, for any $x \in \mathbb{Z}$ and for any finitely supported function $f : \mathbb{Z} \to \mathbb{C}$, we may write

$$\tilde{A}^P_{M_1, M_2; \mathbb{Z}, S_1} f(x) = \mathbb{E}_{m \in R_{M_1, M_2}} f(x - P(m_1, m_2)). \tag{3.18}$$

We shall also abbreviate $\tilde{A}^P_{M_1, M_2; \mathbb{Z}, S_1}$ to $\tilde{A}^P_{M_1, M_2; \mathbb{Z}}$. In fact, we will be able to deduce Theorem 3.16 from its integer counterpart.

Theorem 3.19. *Let $P \in \mathbb{Z}[m_1, m_2]$ be a non-degenerate polynomial (see (1.14)) such that $P(0, 0) = 0$. Let $\tilde{A}^P_{M_1, M_2; \mathbb{Z}} f$ be the average defined in (3.18). If $1 < p < \infty$ and $\tau > 1$, and $\mathbb{D}_\tau := \{\tau^n : n \in \mathbb{N}\}$, then one has*

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}^2_\tau)} \|O_{I, J}(\tilde{A}^P_{M_1, M_2; \mathbb{Z}} f : M_1, M_2 \in \mathbb{D}_\tau)\|_{\ell^P(\mathbb{Z})} \lesssim_{p, \tau, P} \|f\|_{\ell^P(\mathbb{Z})}. \tag{3.20}$$

The implicit constant in (3.20) can be taken to depend only on p, τ, P.

We immediately see that Theorem 3.19 is a special case of Theorem 3.16. However, it is also a standard matter, in view of the Calderón transference principle [19], that this implication can be reversed. So in order to prove (3.17), it suffices to establish (3.20). This reduction is important since we can use Fourier methods in the integer setting which are not readily available in abstract measure spaces.

From now on, we will focus our attention on establishing Theorem 3.19.

4. 'Backwards' Newton diagram: Proof of Theorem 3.19

The 'backwards' Newton diagram N_P of a nontrivial polynomial $P \in \mathbb{R}[m_1, m_2]$,

$$P(m_1, m_2) := \sum_{\gamma_1, \gamma_2} c_{\gamma_1, \gamma_2} m_1^{\gamma_1} m_2^{\gamma_2}, \quad \text{with} \quad c_{0,0} = 0, \tag{4.1}$$

is defined as the closed convex hull of the set

$$\bigcup_{(\gamma_1, \gamma_2) \in S_P} \{(x + \gamma_1, y + \gamma_2) \in \mathbb{R}^2 : x \leq 0, y \leq 0\},$$

where $S_P := \{(\gamma_1, \gamma_2) \in \mathbb{N} \times \mathbb{N} : c_{\gamma_1, \gamma_2} \neq 0\}$ denotes the set of non-vanishing coefficients of P.

Let $V_P \subseteq S_P$ be the set of vertices (corner points) of N_P. Suppose that $V_P := \{v_1, \ldots, v_r\}$, where $v_j = (v_{j,1}, v_{j,2})$ satisfies $v_{j,1} < v_{j+1,1}$, and $v_{j+1,2} < v_{j,2}$ for each $j \in [r]$.

Let $\omega_0 = (0, 1)$ and $\omega_r = (1, 0)$ and for $j \in [r - 1]$, let $\omega_j = (\omega_{j,1}, \omega_{j,2})$ denote a normal vector to the edge $\overline{v_j v_{j+1}} := v_{j+1} - v_j$ such that $\omega_{j,1}, \omega_{j,2}$ are positive integers (the choice is not unique but it is not an issue here). Observe that the slopes of the lines along ω_j's are decreasing as j increases since N_P is convex. The convexity of N_P also yields that

$$\omega_j \cdot (v - v_j) \leq 0 \quad \text{and} \quad \omega_{j-1} \cdot (v - v_j) \leq 0 \quad \text{(with one inequality strict),} \tag{4.2}$$

for all $v \in S_P \setminus \{v_j\}$ and $j \in [r]$. Now for $j \in [r]$, let us define

$$W(j) := \bigcap_{v \in S_P \setminus \{v_j\}} \{(a, b) \in \mathbb{Z}_+ \times \mathbb{Z}_+ : (a, b) \cdot (v - v_j) < 0\},$$

which is the intersection of various half planes. If $V_P = \{v_1\}$, then we simply define $W(1) = \mathbb{Z}_+ \times \mathbb{Z}_+$.

Remark 4.3. Obviously, if $1 \le i < j \le r$, then $W(i) \cap W(j) = \emptyset$. Indeed, if $(a,b) \in W(i) \cap W(j)$, then $(a,b) \cdot (v - v_i) < 0$ for all $v \in S_P \setminus \{v_i\}$ and $(a,b) \cdot (v - v_j) < 0$ for all $v \in S_P \setminus \{v_j\}$. In particular, $(a,b) \cdot (v_j - v_i) < 0$ and $(a,b) \cdot (v_i - v_j) < 0$, which is impossible.

Lemma 4.4. *For $j \in [r]$, we have*

$$W(j) = \{(a,b) \in \mathbb{Z}_+ \times \mathbb{Z}_+ : \exists_{\alpha,\beta>0} \, (a,b) = \alpha\omega_{j-1} + \beta\omega_j\}.$$

Proof. The convexity of N_P implies that the normals ω_{j-1}, ω_j are linearly independent; therefore, for every $(a,b) \in \mathbb{Z}_+ \times \mathbb{Z}_+$, there are α, β such that $(a,b) = \alpha\omega_{j-1} + \beta\omega_j$. We only need to show that $(a,b) \in W(j)$ if and only if $\alpha, \beta > 0$. First, suppose that $(a,b) \in W(j)$. Then $(a,b) \cdot (v - v_j) < 0$ for all $v \in S_P \setminus \{v_j\}$. In particular, $(a,b) \cdot (v_{j+1} - v_j) = (\alpha\omega_{j-1} + \beta\omega_j) \cdot (v_{j+1} - v_j) < 0$. But this implies that $\alpha\omega_{j-1} \cdot (v_{j+1} - v_j) < 0$, since $\omega_j \cdot (v_{j+1} - v_j) = 0$. This immediately gives that $\alpha > 0$, provided that $j \in [r-1]$, since $\omega_{j-1} \cdot (v_{j+1} - v_j) \le 0$ by (4.2). When $j = r$, then $\alpha > 0$ since $\omega_r = (1,0)$ and $0 < b = (a,b) \cdot (0,1) = (\alpha\omega_{r-1} + \beta\omega_r) \cdot (0,1) = \alpha\omega_{r-1} \cdot (0,1) = \alpha\omega_{r-1,2}$. Similarly, taking $v = v_{j-1}$ for $1 < j \le r$, we obtain $\beta > 0$. When $j = 1$, then $\beta > 0$ because $\omega_0 = (0,1)$ and $0 < a = (a,b) \cdot (1,0) = (\alpha\omega_0 + \beta\omega_1) \cdot (1,0) = \beta\omega_1 \cdot (1,0) = \beta\omega_{1,1}$. Conversely, if $\alpha > 0$ and $\beta > 0$, then for any $v \in S_P \setminus \{v_j\}$, we have $(a,b) \cdot (v - v_j) = \alpha\omega_{j-1} \cdot (v - v_j) + \beta\omega_j \cdot (v - v_j) < 0$, since $\omega_{j-1} \cdot (v - v_j) \le 0$ and $\omega_j \cdot (v - v_j) \le 0$, with at least one inequality strict. $\square$

Lemma 4.4 means that $W(j)$ consists of those lattice points of $\mathbb{Z}_+ \times \mathbb{Z}_+$ which are within the cone centered at the origin with the boundaries determined by the lines along the normals ω_{j-1} and ω_j, respectively. Now for $j \in [r]$, we set

$$S(j) := \{(a,b) \in \mathbb{N} \times \mathbb{N} : \exists_{\alpha \ge 0, \beta \ge 0} \, (a,b) = \alpha\omega_{j-1} + \beta\omega_j\}.$$

Remark 4.5. Some comments are in order.

1. Having defined the sets $S(j)$ for $j \in [r]$, it is not difficult to see that

$$\bigcup_{j=1}^{r} S(j) = \mathbb{N} \times \mathbb{N}. \tag{4.6}$$

2. We note that for $(a,b) \in S(j)$, we have $(a,b) \cdot (v - v_j) \le 0$ for all $v \in S_P$ by (4.2). However, the strict inequality may not be achieved even for $v \ne v_j$.
3. If $r \ge 2$, then by construction of the sets $S(j)$, one sees that if $(a,b) \in S(j)$, then

$$\frac{\omega_{j,2}}{\omega_{j,1}} a \le b \le \frac{\omega_{j-1,2}}{\omega_{j-1,1}} a \tag{4.7}$$

for any $1 < j < r$; and if $j = 1$ or $j = r$ one has, respectively,

$$\frac{\omega_{1,2}}{\omega_{1,1}} a \le b < \infty, \qquad \text{and} \qquad 0 \le b \le \frac{\omega_{r-1,2}}{\omega_{r-1,1}} a. \tag{4.8}$$

4. If $r = 1$ and $(a,b) \in S(1)$, then $0 \le a, b < \infty$.

Now for any given $(a,b) \in S(j)$, we try to determine α and β explicitly. Let $A_j := [\omega_{j-1}|\omega_j]$ be the matrix whose column vectors are the normals ω_{j-1}, ω_j. Then

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \omega_{j-1,1} & \omega_{j,1} \\ \omega_{j-1,2} & \omega_{j,2} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

The convexity of N_P (and the orientation we chose) ensures that $\det A_j < 0$. Taking $d_j := -\det A_j > 0$, one has

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \frac{1}{\det A_j} \begin{pmatrix} \omega_{j,2} & -\omega_{j,1} \\ -\omega_{j-1,2} & \omega_{j-1,1} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{d_j} \begin{pmatrix} -a\omega_{j,2} + b\omega_{j,1} \\ a\omega_{j-1,2} - b\omega_{j-1,1} \end{pmatrix}.$$

We have chosen the components of ω_{j-1} and ω_j to be non-negative integers; therefore, for $j \in [r-1]$ (keeping in mind that $\alpha, \beta \geq 0$ and $d_j > 0$), we may rewrite

$$S(j) = \{(a,b) \in \mathbb{N} \times \mathbb{N} : \exists_{(t_1,t_2) \in \mathbb{N} \times \mathbb{N}} \, (a,b) = \frac{t_1}{d_j}\omega_{j-1} + \frac{t_2}{d_j}\omega_j\}.$$

We allow t_1 to be zero when $j = r$.

We now split $S(j)$ into $S_1(j)$ and $S_2(j)$, where

$$S_1(j) := \{(a,b) \in S(j) : (a,b) = \frac{(n+N)}{d_j}\omega_{j-1} + \frac{N}{d_j}\omega_j, \, n \in \mathbb{N}, \, N \in \mathbb{N}\},$$

$$S_2(j) := \{(a,b) \in S(j) : (a,b) = \frac{N}{d_j}\omega_{j-1} + \frac{(n+N)}{d_j}\omega_j, \, n \in \mathbb{N}, \, N \in \mathbb{N}\}.$$

We can further decompose

$$S_1(j) = \bigcup_{N \in \mathbb{N}} S_1^N(j), \qquad \text{and} \qquad S_2(j) = \bigcup_{N \in \mathbb{N}} S_2^N(j),$$

where

$$\begin{aligned} S_1^N(j) &:= \{(a,b) \in S(j) : (a,b) = \frac{(n+N)}{d_j}\omega_{j-1} + \frac{N}{d_j}\omega_j, \, n \in \mathbb{N}\}, \\ S_2^N(j) &:= \{(a,b) \in S(j) : (a,b) = \frac{N}{d_j}\omega_{j-1} + \frac{(n+N)}{d_j}\omega_j, \, n \in \mathbb{N}\}. \end{aligned} \tag{4.9}$$

Lemma 4.10. *For each $j \in [r]$, there exists $\sigma_j > 0$ such that for every $v \in S_P \setminus \{v_j\}$, one has*

$$(a,b) \cdot (v - v_j) \leq -\sigma_j N \tag{4.11}$$

for all $(a,b) \in S_1^N(j)$. The same conclusion is true for $S_2^N(j)$.

Proof. For every $(a,b) \in S_1^N(j)$, we can write

$$(a,b) = \frac{(n+N)}{d_j}\omega_{j-1} + \frac{N}{d_j}\omega_j = \frac{n}{d_j}\omega_{j-1} + \frac{N}{d_j}(\omega_j + \omega_{j-1})$$

for some $n \in \mathbb{N}$. By (4.2), we have

$$(v - v_j) \cdot (\omega_j + \omega_{j-1}) < 0$$

for all $v \in S_P \setminus \{v_j\}$, since ω_{j-1} and ω_j are linearly independent. Taking

$$\sigma_j := \min_{v \in S_P \setminus \{v_j\}} \frac{1}{d_j}(v_j - v) \cdot (\omega_j + \omega_{j-1}) > 0,$$

one sees, by (4.2) again, that

$$(a, b) \cdot (v - v_j) = \frac{n}{d_j} \omega_{j-1} \cdot (v - v_j) + \frac{N}{d_j} (\omega_j + \omega_{j-1}) \cdot (v - v_j) \le -\sigma_j N$$

for all $(a, b) \in S_1^N(j)$. This immediately yields (4.11) and the proof is finished. □

For any $\tau > 1$ using the decomposition (4.6), we may write

$$\mathbb{D}_\tau \times \mathbb{D}_\tau = \bigcup_{j=1}^r \mathbb{S}_\tau(j), \tag{4.12}$$

where

$$\mathbb{S}_\tau(j) := \{(\tau^{n_1}, \tau^{n_2}) \in \mathbb{D}_\tau \times \mathbb{D}_\tau : (n_1, n_2) \in S(j)\}, \quad \text{for} \quad j \in [r]. \tag{4.13}$$

Using (4.9), we can further write

$$\mathbb{S}_\tau(j) = \bigcup_{N \in \mathbb{N}} \mathbb{S}_{\tau,1}^N(j) \cup \bigcup_{N \in \mathbb{N}} \mathbb{S}_{\tau,2}^N(j), \tag{4.14}$$

where for any $j \in [r]$, one has

$$\begin{aligned}
\mathbb{S}_{\tau,1}^N(j) &:= \{(\tau^{n_1}, \tau^{n_2}) \in \mathbb{D}_\tau \times \mathbb{D}_\tau : (n_1, n_2) \in S_1^N(j)\}, \\
\mathbb{S}_{\tau,2}^N(j) &:= \{(\tau^{n_1}, \tau^{n_2}) \in \mathbb{D}_\tau \times \mathbb{D}_\tau : (n_1, n_2) \in S_2^N(j)\}.
\end{aligned} \tag{4.15}$$

In view of decomposition (4.12), our aim will be to restrict the estimates for oscillations to sectors from (4.13).

Theorem 4.16. *Let $P \in \mathbb{Z}[\mathrm{m}_1, \mathrm{m}_2]$ be a non-degenerate polynomial (see (1.14)) such that $P(0, 0) = 0$. Let $r \in \mathbb{Z}_+$ be the number of corners in the corresponding Newton diagram N_P. Let $f \in \ell^p(\mathbb{Z})$ for some $1 \le p \le \infty$, and let $\tilde{A}_{M_1,M_2;\mathbb{Z}}^P f$ be the average defined in (3.18). If $1 < p < \infty$ and $\tau > 1$ and $j \in [r]$, and $\mathbb{S}_\tau(j)$ is a sector from (4.13), then one has*

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{S}_\tau(j))} \left\| O_{I,J}(\tilde{A}_{M_1,M_2;\mathbb{Z}}^P f : (M_1, M_2) \in \mathbb{S}_\tau(j)) \right\|_{\ell^p(\mathbb{Z})} \lesssim_{p,\tau,P} \|f\|_{\ell^p(\mathbb{Z})}. \tag{4.17}$$

The implicit constant in (4.17) may only depend on p, τ, P.

The proof of Theorem 4.16 is postponed to Section 7. However, assuming momentarily Theorem 4.16, we can derive Theorem 3.19.

Proof of Theorem 3.19. Assume that (4.17) holds for all $j \in [r]$. By (4.12) and (2.11), one has

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^2)} \left\| O_{I,J}(\tilde{A}_{M;\mathbb{Z}}^P f : M \in \mathbb{D}_\tau^2) \right\|_{\ell^p(\mathbb{Z})} \lesssim \sum_{j \in [r]} \sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau^2)} \left\| O_{I,J}(\tilde{A}_{M;\mathbb{Z}}^P f : M \in \mathbb{S}_\tau(j)) \right\|_{\ell^p(\mathbb{Z})}.$$

Step 1. If suffices to show that for every $j \in [r]$, every $J \in \mathbb{Z}_+$ and every $I \in \mathfrak{S}_J(\mathbb{D}_\tau^2)$, one has

$$\left\| \left(\sum_{i \in \mathbb{N}_{<J}} \sup_{M \in \mathbb{B}[I,i] \cap \mathbb{S}_\tau(j)} |\tilde{A}_{M;\mathbb{Z}}^P f - \tilde{A}_{I_i;\mathbb{Z}}^P f|^2 \right)^{1/2} \right\|_{\ell^p(\mathbb{Z})}$$

$$\lesssim \sum_{j \in [r]} \sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{S}_\tau(j))} \left\| O_{I,J}(\tilde{A}_{M;\mathbb{Z}}^P f : M \in \mathbb{S}_\tau(j)) \right\|_{\ell^p(\mathbb{Z})}. \tag{4.18}$$

We can assume that $J > Cr$ for a large $C > 0$; otherwise, the estimate in (4.18) easily follows from maximal function estimates. Let us fix a sequence $I = (I_0, \ldots, I_J) \in \mathfrak{S}_J(\mathbb{D}_\tau^2)$ and a sector $\mathbb{S}_\tau(j)$. Let $\omega_* := \max\{\omega_{i1}, \omega_{i2} : i \in [r]\}$ and we split the set $\mathbb{N}_{<J}$ into $O(r)$ sparse sets $\mathbb{J}_1, \ldots, \mathbb{J}_{O(r)} \subset \mathbb{N}_{<J}$, where each $\mathbb{J} \in \{\mathbb{J}_1, \ldots, \mathbb{J}_{O(r)}\}$ satisfies the separation condition:

$$\log_\tau I_{i_2 1} - \log_\tau I_{(i_1+1)1} \geq 100r\omega_* \quad \text{and} \quad \log_\tau I_{i_2 2} - \log_\tau I_{(i_1+1)2} \geq 100r\omega_* \quad (4.19)$$

for every $i_1, i_2 \in \mathbb{J}$ such that $i_1 < i_2$. Our task now is to establish (4.18) with the summation over $\mathbb{J}$ satisfying (4.19) in place of $\mathbb{N}_{<J}$ in the sum on the left-hand side of (4.18).

Step 2. To every element $I_i = (I_{i1}, I_{i2})$ with $i \in \mathbb{N}_{<J}$ in the sequence I (which say lies in the sector $\mathbb{S}_\tau(j_i)$), we associate at most one point $P_i(j) \in \mathbb{S}_\tau(j)$ in the following way. If $j_i < j$ and the box $\mathbb{B}[I, i]$ intersects the sector $\mathbb{S}_\tau(j)$, then the box intersects the sector along the bottom edge. We set $P_i(j) = (I_i^j, I_{i2})$, where I_i^j is the least element in $\mathbb{D}_\tau$ such that $(I_i^j, I_{i2}) \in \mathbb{S}_\tau(j)$. If $j < j_i$ and the box $\mathbb{B}[I, i]$ intersects the sector $\mathbb{S}_\tau(j)$, then it intersects the sector along the left edge. We set $P_i(j) = (I_{i1}, \tilde{I}_i^j)$, where $\tilde{I}_i^j$ is the least element in $\mathbb{D}_\tau$ such that $(I_{i1}, \tilde{I}_i^j) \in \mathbb{S}_\tau(j)$. Finally if $j_i = j$, we set $P_i(j) = I_i$. The sequence $P(j) = (P_i(j) : i \in \mathbb{N}_{\leq J'})$ forms a strictly increasing sequence lying in $\mathfrak{S}_{J'}(\mathbb{S}_\tau(j))$ for some $J' \leq J$ and each $P_i(j) = (P_{i1}(j), P_{i2}(j))$ is the least element among all the elements $(M_1, M_2) \in \mathbb{B}[I, i] \cap \mathbb{S}_\tau(j)$.

Step 3. We now produce a sequence of length at most $r+2$, which will allow us to move from I_i to $P_i(j)$ when $I_i \neq P_i(j)$. More precisely, we claim that there exists a sequence $u^i := (u_m^i : m \in \mathbb{N}_{<m_{I_i}}) \subset \mathbb{D}_\tau^2$ for some $m_{I_i} \in [r+1]$, with the property that

$$u_0^i > u_1^i > \ldots > u_{m_{I_i}-1}^i, \quad \text{and} \quad u_{m_{I_i}-1}^i \prec u_{m_{I_i}}^i, \quad (4.20)$$

where $(u_0^i, u_{m_{I_i}}^i) = (I_i, P_i(j))$ or $(u_0^i, u_{m_{I_i}}^i) = (P_i(j), I_i)$. Moreover, two consecutive elements u_m^i, u_{m+1}^i of this sequence belong to a unique sector $\mathbb{S}_\tau(j_{u_m^i})$ except the elements $u_{m_{I_i}-2}^i, u_{m_{I_i}-1}^i$ and $u_{m_{I_i}-1}^i, u_{m_{I_i}}^i$, which may belong to the same sector. Suppose now that $\mathbb{B}[I, i] \cap \mathbb{S}_\tau(j) \neq \emptyset$ and $I_i \in \mathbb{S}_\tau(j_i)$ and $j_i < j$. Let $u_0^i := I_i$ be the starting point. Suppose that the elements $u_0^i > u_1^i > \ldots > u_m^i$ have been chosen for some $m \in \mathbb{N}_{<r}$ so that u_s^i lies on the bottom boundary ray of $\mathbb{S}_\tau(j_i + s - 1)$ and $u_s^i \prec u_{s-1}^i$ for each $s \in [m]$. Then we take u_m^i and move southwesterly to u_{m+1}^i, the nearest point on the bottom boundary ray of $\mathbb{S}_\tau(j_i+m)$ such that $u_{m+1}^i \prec u_m^i$. Continuing this way after $m_{I_i} - 1 = j - j_i + 1 \leq r$ steps, we arrive at $u_{m_{I_i}-1}^i \in \mathbb{S}_\tau(j)$ which will allow us to reach the last point of this sequence $u_{m_{I_i}}^i := P_i(j)$ as claimed in (4.20). Assume now that $\mathbb{B}[I, i] \cap \mathbb{S}_\tau(j) \neq \emptyset$ and $I_i \in \mathbb{S}_\tau(j_i)$ and $j_i > j$. We start from the point $u_0^i := P_i(j)$ and proceed exactly the same as in the previous case until we reach the point $u_{m_{I_i}}^i := I_i$.

Step 4. To complete the proof, we use the sequence from (4.20) for each $i \in \mathbb{J}$ and observe that

$$\left\| \left(\sum_{i \in \mathbb{J}} \sup_{M \in \mathbb{B}[I,i] \cap \mathbb{S}_\tau(j)} |\tilde{A}_{M;\mathbb{Z}}^P f - \tilde{A}_{I_i;\mathbb{Z}}^P f|^2 \right)^{1/2} \right\|_{\ell^p(\mathbb{Z})}$$

$$\leq \left\| \left(\sum_{i \in \mathbb{J}} \sup_{M \in \mathbb{B}[P(j),i] \cap \mathbb{S}_\tau(j)} |\tilde{A}_{M;\mathbb{Z}}^P f - \tilde{A}_{I_i;\mathbb{Z}}^P f|^2 \right)^{1/2} \right\|_{\ell^p(\mathbb{Z})}$$

$$\lesssim_r \left\| \left(\sum_{i \in \mathbb{J}} \sup_{M \in \mathbb{B}[P(j),i] \cap \mathbb{S}_\tau(j)} |\tilde{A}_{M;\mathbb{Z}}^P f - \tilde{A}_{P_i(j);\mathbb{Z}}^P f|^2 \right)^{1/2} \right\|_{\ell^p(\mathbb{Z})}$$

$$+ \left\| \left(\sum_{i \in \mathbb{J}} \sum_{m \in \mathbb{N}_{<m_{I_i}}} |\tilde{A}_{u_{m+1}^i;\mathbb{Z}}^P f - \tilde{A}_{u_m^i;\mathbb{Z}}^P f|^2 \right)^{1/2} \right\|_{\ell^p(\mathbb{Z})}.$$

Clearly, the first norm is dominated by the right-hand side of (4.18). The same is true for the second norm. It follows from the fact that for two consecutive integers $i_1 < i_2$ such that $\mathbb{B}[I, i_1] \cap \mathbb{S}_\tau(j) \neq \emptyset$ and $\mathbb{B}[I, i_2] \cap \mathbb{S}_\tau(j) \neq \emptyset$, if we have $u_{j_1}^{i_1}$ and $u_{j_2}^{i_2}$ belonging to the same sector, they must satisfy $u_{j_1}^{i_1} \prec u_{j_2}^{i_2}$ by the separation condition (4.19). This completes the proof of the theorem. □

5. Exponential sum estimates

This section is intended to establish certain double exponential sum estimates which will be used later. We begin by recalling the classical Weyl inequality with a logarithmic loss.

Proposition 5.1. *Let $d \in \mathbb{Z}_+$, $d \geq 2$ and let $P \in \mathbb{R}[m]$ be such that $P(m) := c_d m^d + \ldots + c_1 m$. Then there exists a constant $C > 0$ such that for every $M \in \mathbb{Z}_+$, the following is true. Suppose that for some $2 \leq j \leq d$, there are $a, q \in \mathbb{Z}$ such that $1 \leq q \leq M^j$ and $(a, q) = 1$ and*

$$\left| c_j - \frac{a}{q} \right| \leq \frac{1}{q^2}.$$

Then for $\sigma(d) := 2d^2 - 2d + 1$, one has

$$\left| \sum_{m=1}^{M} e(P(m)) \right| \leq CM \log(2M) \left(\frac{1}{q} + \frac{1}{M} + \frac{q}{M^j} \right)^{\frac{1}{\sigma(d)}}. \tag{5.2}$$

For the proof, we refer to [70, Theorem 1.5]. The range of summation in (5.2) can be shifted to any segment of length M without affecting the bound. We will also recall a simple lemma from [52, Lemma A.15, p. 53] (see also [58, Lemma 1, p. 1298]), which follows from the Dirichlet principle.

Lemma 5.3. *Let $\theta \in \mathbb{R}$ and $Q \in \mathbb{Z} \setminus \{0\}$. Suppose that*

$$\left| \theta - \frac{a}{q} \right| \leq \frac{1}{q^2}$$

for some integers $0 \leq a < q \leq M$ with $(a, q) = 1$ for some $M \geq 1$. Then there is a reduced fraction a'/q' so that $(a', q') = 1$ and

$$\left| Q\theta - \frac{a'}{q'} \right| \leq \frac{1}{2q'M}$$

with $q/(2|Q|) \leq q' \leq 2M$.

We now extend Weyl's inequality in Proposition 5.1 to include the $j = 1$ case.

Proposition 5.4. *Let $d \in \mathbb{Z}_+$ and let $P \in \mathbb{R}[m]$ be such that $P(m) := c_d m^d + \ldots + c_1 m$. Then there exists a constant $C > 0$ such that for every $M \in \mathbb{Z}_+$, the following is true. Suppose that for some $1 \leq j \leq d$, there are $a, q \in \mathbb{Z}$ such that $1 \leq q \leq M^j$ and $(a, q) = 1$ and*

$$\left| c_j - \frac{a}{q} \right| \leq \frac{1}{q^2}. \tag{5.5}$$

Then for certain $\tau(d) \in \mathbb{Z}_+$, one has

$$\left| \sum_{m=1}^{M} e(P(m)) \right| \leq CM \log(2M) \left(\frac{1}{q} + \frac{1}{M} + \frac{q}{M^j} \right)^{\frac{1}{\tau(d)}}. \tag{5.6}$$

Proof. We first assume that $d = 1$. Then $P(m) = c_1 m$ and $j = 1$. We can also assume that $q \geq 2$; otherwise, (5.6) is obvious. Now it is easy to see that

$$\left| \sum_{m=1}^{M} e(c_1 m) \right| \leq \frac{1}{\|c_1\|} \lesssim q.$$

Thus, (5.6) holds with $\tau(1) = 1$. Now we assume that $d \geq 2$. If (5.5) holds for some $2 \leq j \leq d$, then (5.6) follows from Proposition 5.1 with $\tau(d) = \sigma(d)$, where $\sigma(d)$ is the exponent as in (5.2). Hence,

we can assume that $j = 1$. Define $\kappa := \min\{q, M/q\}$, and let $\chi \in (0, (4d)^{-1})$. We may assume that $\kappa > 100$; otherwise, (5.6) obviously follows. For every $2 \leq j' \leq d$, by Dirichlet's principle, there is a reduced fraction $a_{j'}/q_{j'}$ such that

$$\left| c_{j'} - \frac{a_{j'}}{q_{j'}} \right| \leq \frac{\kappa^\chi}{q_{j'} M^{j'}} \tag{5.7}$$

with $(a_{j'}, q_{j'})$ and $1 \leq q'_j \leq M^{j'} \kappa^{-\chi}$. We may assume that $1 \leq q_{j'} \leq \kappa^\chi$ for all $2 \leq j' \leq d$, since otherwise the claim follows from (5.2) with $\tau(d) = \lceil \sigma(d)\chi^{-1} \rceil$. Let $Q := \mathrm{lcm}\{q_{j'} : 2 \leq j' \leq d\} \leq \kappa^{d\chi}$ and note that $Q \leq M$ follows from the definition of κ. We have

$$\left| \sum_{m=1}^{M} e(P(m)) \right| \leq \sum_{r=1}^{Q} \left| \sum_{-\frac{r}{Q} < \ell \leq \frac{M-r}{Q}} e(P(Q\ell + r)) \right|$$

$$= \sum_{r=1}^{Q} \left| \sum_{U < \ell \leq V} A_\ell B_\ell \right|,$$

where $U := -\frac{r}{Q}$, $V := \frac{M-r}{Q}$ and $A_\ell := e(c_1 Q\ell)$ and

$$B_\ell := e\left(\sum_{j'=2}^{d} c_{j'}(Q\ell + r)^{j'} \right) = e\left(\sum_{j'=2}^{d} \alpha_{j'}(Q\ell + r)^{j'} + \sum_{j'=2}^{d} \frac{a_{j'}}{q_{j'}} r^{j'} \right),$$

where $\alpha_{j'} := c_{j'} - a_{j'}/q_{j'}$ satisfies the estimate (5.7). Using the summation by parts formula (2.1), we obtain

$$\sum_{U < \ell \leq V} A_\ell B_\ell = S_V B_{\lfloor V \rfloor} + \sum_{\ell \in (U, V-1] \cap \mathbb{Z}} S_\ell (B_\ell - B_{\ell+1}),$$

with $S_\ell := \sum_{k \in (U, \ell] \cap \mathbb{Z}} A_k$.

From above, since $Q \leq M$, we see that

$$|B_{\ell+1} - B_\ell| \lesssim \kappa^\chi Q M^{-1}.$$

By Lemma 5.3 (with $M = q$), there is a reduced fraction a'/q' such that $(a', q') = 1$ and

$$\left| c_1 Q - \frac{a'}{q'} \right| \leq \frac{1}{2qq'} \qquad \text{and} \qquad \kappa^{1-d\chi}/2 \leq q' \leq 2q \leq 2M/\kappa.$$

Hence, $q' \geq \kappa^{1-d\chi}/2 \geq 2$ and so

$$|S_\ell| \lesssim \frac{1}{\|c_1 Q\|} \lesssim q' \lesssim M/\kappa.$$

Consequently, we conclude that

$$\left| \sum_{m=1}^{M} e(P(m)) \right| \lesssim M\kappa^{-1/2}.$$

This implies (5.6) with $\tau(d) = 2$, and the proof of Proposition 5.4 is complete.   □

We shall also use the Vinogradov mean value theorem. A detailed exposition of Vinogradov's method can be found in [35, Section 8.5, p. 216]; see also [70]. We shall follow [35]. For each integer $s \geq 1$ and

$k, N \geq 2$ and for $\lambda_1, \ldots, \lambda_k \in \mathbb{Z}$ let $J_{s,k}(N; \lambda_1, \ldots, \lambda_k)$ denote the number of solutions to the system of k inhomogeneous equations in $2s$ variables given by

$$
\begin{cases}
x_1 + \ldots + x_s - y_1 - \ldots - y_s = \lambda_1 \\
x_1^2 + \ldots + x_s^2 - y_1^2 - \ldots - y_s^2 = \lambda_2 \\
\qquad\qquad\quad\vdots \\
x_1^k + \ldots + x_s^k - y_1^k - \ldots - y_s^k = \lambda_k,
\end{cases}
\tag{5.8}
$$

where $x_j, y_j \in [N]$ for every $j \in [s]$. The number $J_{s,k}(N; \lambda_1, \ldots, \lambda_k)$ can be expressed in terms of a certain exponential sum. Let $R_k(x) := (x, x^2, \ldots, x^k) \in \mathbb{R}^k$ denote the moment curve for $x \in \mathbb{R}$. For $\xi = (\xi_1, \ldots, \xi_k) \in \mathbb{R}^k$, define the exponential sum

$$
S_k(\xi; N) := \sum_{n=1}^{N} e(\xi \cdot R_k(n)) = \sum_{n=1}^{N} e(\xi_1 n + \ldots + \xi_k n^k).
$$

One easily obtains

$$
|S_k(\xi; N)|^{2s} = \sum_{|\lambda_1| \leq sN} \cdots \sum_{|\lambda_k| \leq sN^k} J_{s,k}(N; \lambda_1, \ldots, \lambda_k) e(\xi \cdot \lambda),
\tag{5.9}
$$

which by the Fourier inversion formula gives

$$
J_{s,k}(N; \lambda_1, \ldots, \lambda_k) = \int_{[0,1)^k} |S_k(\xi; N)|^{2s} e(-\xi \cdot \lambda) d\xi.
\tag{5.10}
$$

Moreover, from (5.10), one has

$$
J_{s,k}(N; \lambda_1, \ldots, \lambda_k) \leq J_{s,k}(N) := J_{s,k}(N; 0, \ldots, 0),
\tag{5.11}
$$

where the number $J_{s,k}(N)$ represents the number of solutions to the system of k homogeneous equations in $2s$ variables as in (5.8) with $\lambda_1 = \ldots = \lambda_k = 0$.

Vinogradov's mean value theorem can be formulated as follows:

Theorem 5.12. *For all integers $s \geq 1$ and $k \geq 2$ and any $\varepsilon > 0$, there is a constant $C_\varepsilon > 0$ such that for every integer $N \geq 2$, one has*

$$
J_{s,k}(N) \leq C_\varepsilon \left(N^{s+\varepsilon} + N^{2s - \frac{k(k+1)}{2} + \varepsilon} \right).
\tag{5.13}
$$

Moreover, if additionally $s > \frac{1}{2} k(k+1)$, then there is a constant $C > 0$ such that

$$
J_{s,k}(N) \leq C N^{2s - \frac{k(k+1)}{2}}.
\tag{5.14}
$$

Apart from the N^ε loss in (5.13), this bound is known to be sharp. Inequality (5.13) is fairly simple for $k = 2$ and follows from elementary estimates for the divisor function. The conclusion of Theorem 5.12 for $k \geq 3$, known as Vinogradov's mean value theorem, was a central problem in analytic number theory and had been open until recently. The cubic case $k = 3$ was solved by Wooley [69] using the efficient congruencing method. The case for any $k \geq 3$ was solved by the first author with Demeter and Guth [17] using the decoupling method. Not long afterwards, Wooley [68] also showed that the efficient congruencing method can be used to solve the Vinogradov mean value conjecture for all $k \geq 3$. In fact, later we will only use (5.14), which easily follows from (5.13); the details can be found in [17, Section 5].

5.1. Double Weyl's inequality

Let $K_1, K_2 \in \mathbb{N}$, $M_1, M_2 \in \mathbb{Z}_+$ satisfy $K_1 < M_1$ and $K_2 < M_2$. Let $Q \in \mathbb{R}[m_1, m_2]$ be given and define double exponential sums by

$$S_{K_1,M_1,K_2,M_2}(Q) := \sum_{m_1=K_1+1}^{M_1} \sum_{m_2=K_2+1}^{M_2} e(Q(m_1, m_2)), \tag{5.15}$$

$$S^1_{K_1,M_1,K_2,M_2}(Q) := \sum_{m_1=K_1+1}^{M_1} \left| \sum_{m_2=K_2+1}^{M_2} e(Q(m_1, m_2)) \right|, \tag{5.16}$$

$$S^2_{K_1,M_1,K_2,M_2}(Q) := \sum_{m_2=K_2+1}^{M_2} \left| \sum_{m_1=K_1+1}^{M_1} e(Q(m_1, m_2)) \right|. \tag{5.17}$$

If $K_1 = K_2 = 0$, we will abbreviate (5.15), (5.16) and (5.17), respectively, to

$$S_{M_1,M_2}(Q), \qquad S^1_{M_1,M_2}(Q), \quad \text{and} \quad S^2_{M_1,M_2}(Q). \tag{5.18}$$

By the triangle inequality, we have

$$|S_{K_1,M_1,K_2,M_2}(Q)| \le S^1_{K_1,M_1,K_2,M_2}(Q), \quad \text{and} \quad |S_{K_1,M_1,K_2,M_2}(Q)| \le S^2_{K_1,M_1,K_2,M_2}(Q). \tag{5.19}$$

We now provide estimates for (5.15), (5.16) and (5.17) in the spirit of Proposition 5.1 above. We first recall a technical lemma from [39, Chapter IV, Lemma 5, p. 82].

Lemma 5.20. *Let $\alpha \in \mathbb{R}$ and suppose that there are $a \in \mathbb{Z}, q \in \mathbb{Z}_+$ such that $(a, q) = 1$ and*

$$\left| \alpha - \frac{a}{q} \right| \le \frac{1}{q^2}.$$

Then for every $\beta \in \mathbb{R}$, $U > 0$ and $P \ge 1$, one has

$$\sum_{n=1}^{P} \min \left\{ U, \frac{1}{\|\alpha n + \beta\|} \right\} \le 6 \left(1 + \frac{P}{q} \right)(U + q \log q). \tag{5.21}$$

Estimate (5.21) will be useful in the proof of the following counterpart of Weyl's inequality for double sums.

Proposition 5.22. *Let $d_1, d_2 \in \mathbb{Z}_+$ and $Q \in \mathbb{R}[m_1, m_2]$ be such that*

$$Q(m_1, m_2) := \sum_{\gamma_1=0}^{d_1} \sum_{\gamma_2=0}^{d_2} c_{\gamma_1,\gamma_2} m_1^{\gamma_1} m_2^{\gamma_2}, \quad \text{and} \quad c_{0,0} = 0.$$

Then there exists a constant $C > 0$ such that for every $K_1, K_2 \in \mathbb{N}$, $M_1, M_2 \in \mathbb{Z}_+$ satisfying $K_1 \le M_1$ and $K_2 \le M_2$, the following holds. Suppose that for some $1 \le \rho_1 \le d_1$ and $1 \le \rho_2 \le d_2$, there are $a_{\rho_1,\rho_2} \in \mathbb{Z}, q_{\rho_1,\rho_2} \in \mathbb{Z}_+$ such that $(a_{\rho_1,\rho_2}, q_{\rho_1,\rho_2}) = 1$ and

$$\left| c_{\rho_1,\rho_2} - \frac{a_{\rho_1,\rho_2}}{q_{\rho_1,\rho_2}} \right| \le \frac{1}{q_{\rho_1,\rho_2}^2}. \tag{5.23}$$

Set $k_i := d_i(d_i + 1)$ for $i \in [2]$, $M_- := \min(M_1^{\rho_1}, M_2^{\rho_2})$ and $M_+ := \max(M_1^{\rho_1}, M_2^{\rho_2})$. Then for $i \in [2]$,

$$S^i_{K_1, M_1, K_2, M_2}(Q) \le CM_1 M_2 \left(\frac{1}{M_-} + \frac{q_{\rho_1, \rho_2} \log q_{\rho_1, \rho_2}}{M_1^{\rho_1} M_2^{\rho_2}} + \frac{1}{q_{\rho_1, \rho_2}} + \frac{\log q_{\rho_1, \rho_2}}{M_+} \right)^{\frac{1}{4 k_1 k_2}}. \tag{5.24}$$

In view of (5.19), estimates (5.24) clearly hold for $|S_{K_1, M_1, K_2, M_2}(Q)|$.

Remark 5.25. The bracketed expression in (5.24) is equal to $\min(A, B)$ where

$$A = \frac{1}{M_2^{\rho_2}} + \frac{q_{\rho_1, \rho_2} \log q_{\rho_1, \rho_2}}{M_1^{\rho_1} M_2^{\rho_2}} + \frac{1}{q_{\rho_1, \rho_2}} + \frac{\log q_{\rho_1, \rho_2}}{M_1^{\rho_1}}$$

and

$$B = \frac{1}{M_1^{\rho_1}} + \frac{q_{\rho_1, \rho_2} \log q_{\rho_1, \rho_2}}{M_1^{\rho_1} M_2^{\rho_2}} + \frac{1}{q_{\rho_1, \rho_2}} + \frac{\log q_{\rho_1, \rho_2}}{M_2^{\rho_2}}.$$

Multi-parameter exponential sums were extensively investigated over the years. The best source about this subject is [1]. However, here we need bounds as in (5.24), which will allow us to gain logarithmic factors on minor arcs (see Proposition 5.37) in contrast to polynomial factors, which were obtained in [1]. We prove Proposition 5.22 by giving an argument based on an iterative application of the Vinogradov mean value theorem.

Proof of Proposition 5.22. We only prove (5.24) for $i = 1$. The proof of (5.24) for $i = 2$ can be obtained similarly by symmetry. To prove inequality (5.24) when $i = 1$, we shall follow [35, Section 8.5., p. 216] and proceed in five steps.

Step 1. For $i \in [2]$, let us define the d_i-dimensional box

$$\mathcal{B}_{d_i}(M_i) := \left(\prod_{j=1}^{d_i} [-k_i M_i^j, k_i M_i^j] \right) \cap \mathbb{Z}^{d_i}.$$

Observe that

$$Q(m_1, m_2) = \sum_{\gamma_2 = 0}^{d_2} c_{\gamma_2}(m_1) m_2^{\gamma_2} = c(m_1) \cdot R_{d_2}(m_2) + c_0(m_1),$$

where for $\gamma_2 \in [d_2] \cup \{0\}$, one has

$$c(m_1) := (c_1(m_1), \dots, c_{d_2}(m_1)) \qquad \text{and} \qquad c_{\gamma_2}(m_1) := \sum_{\gamma_1 = 0}^{d_1} c_{\gamma_1, \gamma_2} m_1^{\gamma_1}.$$

Recall that $R_{d_2}(m_2) = (m_2, m_2^2, \dots, m_2^{d_2})$. By (5.16), we note that

$$S^1_{K_1, M_1, K_2, M_2}(Q) \le S^1_{M_1, M_2}(Q) + S^1_{M_1, K_2}(Q) \lesssim \max_{N_2 \in [M_2]} S^1_{M_1, N_2}(Q).$$

For any $k_2 \in \mathbb{Z}_+$, by Hölder's inequality and by (5.9), we obtain

$$
\begin{aligned}
S^1_{K_1,M_1,K_2,M_2}(Q)^{2k_2} &\lesssim M_1^{2k_2-1} \max_{N_2 \in [M_2]} \sum_{m_1=1}^{M_1} |S_{d_2}(c(m_1); N_2)|^{2k_2} \\
&= M_1^{2k_2-1} \max_{N_2 \in [M_2]} \sum_{u \in \mathcal{B}_{d_2}(N_2)} J_{k_2,d_2}(N_2; u) \sum_{m_1=1}^{M_1} e(c(m_1) \cdot u).
\end{aligned}
$$

(5.26)

Step 2. We see that

$$
c(m_1) \cdot u = \sum_{\gamma_1=0}^{d_1} \sum_{\gamma_2=1}^{d_2} c_{\gamma_1,\gamma_2} u_{\gamma_2} m_1^{\gamma_1} = \beta^1(u) \cdot R_{d_1}(m_1) + \beta_0^1(u),
$$

where for $u = (u_1, \ldots, u_{d_2}) \in \mathbb{Z}^{d_2}$ and $\gamma_1 \in [d_1] \cup \{0\}$ we set

$$
\beta^1(u) := (\beta_1^1(u), \ldots, \beta_{d_1}^1(u)) \qquad \text{and} \qquad \beta_{\gamma_1}^1(u) := \sum_{\gamma_2=1}^{d_2} c_{\gamma_1,\gamma_2} u_{\gamma_2}.
$$

Similarly, for $v = (v_1, \ldots, v_{d_1}) \in \mathbb{Z}^{d_1}$ and $\gamma_2 \in [d_2] \cup \{0\}$, we also set

$$
\beta^2(v) := (\beta_1^2(v), \ldots, \beta_{d_2}^2(v)) \qquad \text{and} \qquad \beta_{\gamma_2}^2(v) := \sum_{\gamma_1=1}^{d_1} c_{\gamma_1,\gamma_2} v_{\gamma_1}.
$$

This implies, raising both sides of (5.26) to power $2k_1$ for any $k_1 \in \mathbb{Z}_+$, that

$$
\begin{aligned}
S^1_{K_1,M_1,K_2,M_2}(Q)^{4k_1k_2} &\lesssim M_1^{4k_1k_2-2k_1} \max_{N_2 \in [M_2]} \left(\sum_{u \in \mathcal{B}_{d_2}(N_2)} J_{k_2,d_2}(N_2; u) |S_{d_1}(\beta^1(u); M_1)| \right)^{2k_1} \\
&\lesssim M_1^{4k_1k_2-2k_1} M_2^{4k_1k_2-2k_2} \max_{N_2 \in [M_2]} \sum_{u \in \mathcal{B}_{d_2}(N_2)} J_{k_2,d_2}(N_2; u) |S_{d_1}(\beta^1(u); M_1)|^{2k_1}.
\end{aligned}
$$

(5.27)

In (5.27), we used Hölder's inequality and

$$
\sum_{u \in \mathcal{B}_{d_2}(N_2)} J_{k_2,d_2}(N_2; u) = N_2^{2k_2}.
$$

Step 3. For $v = (v_1, \ldots, v_{d_1}) \in \mathbb{Z}^{d_1}$, we have

$$
\beta^1(u) \cdot v = \sum_{\gamma_2=1}^{d_2} \sum_{\gamma_1=1}^{d_1} c_{\gamma_1,\gamma_2} v_{\gamma_1} u_{\gamma_2} = \beta^2(v) \cdot u.
$$

(5.28)

Applying (5.9) and (5.11) to the last sum in (5.27), we obtain

$$
\max_{N_2 \in [M_2]} \sum_{u \in \mathcal{B}_{d_2}(N_2)} J_{k_2,d_2}(N_2; u) |S_{d_1}(\beta^1(u); M_1)|^{2k_1}
$$

$$
\leq J_{k_2,d_2}(M_2) \sum_{u \in \mathcal{B}_{d_2}(M_2)} |S_{d_1}(\beta^1(u); M_1)|^{2k_1}
$$

$$= J_{k_2,d_2}(M_2) \sum_{u \in \mathcal{B}_{d_2}(M_2)} \sum_{v \in \mathcal{B}_{d_1}(M_1)} J_{k_1,d_1}(M_1; v) e(\beta^1(u) \cdot v)$$

$$\leq J_{k_1,d_1}(M_1) J_{k_2,d_2}(M_2) \sum_{v \in \mathcal{B}_{d_1}(M_1)} \Big| \sum_{u \in \mathcal{B}_{d_2}(M_2)} e(\beta^2(v) \cdot u)\Big|, \qquad (5.29)$$

where we used (5.28) in the last inequality. In a slightly more involved process, we now obtain a different estimate for the last sum in (5.27). We apply (5.9) twice to obtain

$$\max_{N_2 \in [M_2]} \sum_{u \in \mathcal{B}_{d_2}(N_2)} J_{k_2,d_2}(N_2; u) |S_{d_1}(\beta^1(u); M_1)|^{2k_1}$$

$$= \max_{N_2 \in [M_2]} \sum_{u \in \mathcal{B}_{d_2}(N_2)} \sum_{v \in \mathcal{B}_{d_1}(M_1)} J_{k_2,d_2}(N_2; u) J_{k_1,d_1}(M_1; v) e(\beta^1(u) \cdot v)$$

$$= \max_{N_2 \in [M_2]} \sum_{v \in \mathcal{B}_{d_1}(M_1)} J_{k_1,d_1}(M_1; v) \sum_{u \in \mathcal{B}_{d_2}(N_2)} J_{k_2,d_2}(N_2; u) e(\beta^2(v) \cdot u)$$

$$= \max_{N_2 \in [M_2]} \sum_{v \in \mathcal{B}_{d_1}(M_1)} J_{k_1,d_1}(M_1; v) |S_{d_2}(\beta^2(v); N_2)|^{2k_2},$$

where we used (5.28) in the penultimate equality. Hence, by (5.9), (5.11) and (5.28),

$$\max_{N_2 \in [M_2]} \sum_{u \in \mathcal{B}_{d_2}(N_2)} J_{k_2,d_2}(N_2; u) |S_{d_1}(\beta^1(u); M_1)|^{2k_1}$$

$$\leq J_{k_1,d_1}(M_1) \max_{N_2 \in [M_2]} \sum_{v \in \mathcal{B}_{d_1}(M_1)} |S_{d_2}(\beta^2(v); N_2)|^{2k_2}$$

$$= J_{k_1,d_1}(M_1) \max_{N_2 \in [M_2]} \sum_{u \in \mathcal{B}_{d_2}(N_2)} J_{k_2,d_2}(N_2; u) \sum_{v \in \mathcal{B}_{d_1}(M_1)} e(\beta^1(u) \cdot v)$$

$$\leq J_{k_1,d_1}(M_1) J_{k_2,d_2}(M_2) \sum_{u \in \mathcal{B}_{d_2}(M_2)} \Big| \sum_{v \in \mathcal{B}_{d_1}(M_1)} e(\beta^1(u) \cdot v)\Big|. \qquad (5.30)$$

Step 4. In this step, we prove (for $q = q_{\rho_1,\rho_2}$)

$$\sum_{u \in \mathcal{B}_{d_2}(M_2)} \Big| \sum_{v \in \mathcal{B}_{d_1}(M_1)} e(\beta^1(u) \cdot v)\Big| \lesssim \prod_{j=1}^{2} M_j^{\frac{d_j(d_j+1)}{2}} \left(\frac{1}{M_2^{\rho_2}} + \frac{q \log q}{M_1^{\rho_1} M_2^{\rho_2}} + \frac{1}{q} + \frac{\log q}{M_1^{\rho_1}} \right) \qquad (5.31)$$

and

$$\sum_{v \in \mathcal{B}_{d_1}(M_1)} \Big| \sum_{u \in \mathcal{B}_{d_2}(M_2)} e(\beta^2(v) \cdot u)\Big| \lesssim \prod_{j=1}^{2} M_j^{\frac{d_j(d_j+1)}{2}} \left(\frac{1}{M_1^{\rho_1}} + \frac{q \log q}{M_1^{\rho_1} M_2^{\rho_2}} + \frac{1}{q} + \frac{\log q}{M_2^{\rho_2}} \right). \qquad (5.32)$$

We only establish (5.31). The symmetric bound (5.32) is similar. The exponential sum

$$\sum_{v \in \mathcal{B}_{d_1}(M_1)} e(\beta^1(u) \cdot v) = \prod_{\gamma_1=1}^{d_1} \sum_{|v_{\gamma_1}| \leq k_1 M_1^{\gamma_1}} e(\beta_{\gamma_1}^1(u) v_{\gamma_1})$$

is a product of geometric series which we can easily evaluate to conclude

$$\sum_{u \in \mathcal{B}_{d_2}(M_2)} \Big| \sum_{v \in \mathcal{B}_{d_1}(M_1)} \boldsymbol{e}(\beta^1(u) \cdot v) \Big| \lesssim \sum_{u \in \mathcal{B}_{d_2}(M_2)} \prod_{\gamma_1=1}^{d_1} \min \Big\{ 2d_1 M_1^{\gamma_1}, \frac{1}{\|\beta^1_{\gamma_1}(u)\|} \Big\}$$

$$\leq (2d_1 M_1)^{\frac{d_1(d_1+1)}{2}-\rho_1} \sum_{u \in \mathcal{B}_{d_2}(M_2)} \min \Big\{ 2d_1 M_1^{\rho_1}, \frac{1}{\|\beta^1_{\rho_1}(u)\|} \Big\}.$$

Since (5.23) holds and

$$\beta^1_{\rho_1}(u) = c_{\rho_1,\rho_2} u_{\rho_2} + \beta(u), \qquad \text{where} \qquad \beta(u) := \sum_{\substack{\gamma_2=1 \\ \gamma_2 \neq \rho_2}}^{d_2} c_{\rho_1,\gamma_2} u_{\gamma_2},$$

we can apply (5.21) with $P = k_2 M_2^{\rho_2}$, $U = 2d_1 M_1^{\rho_1}$ and $q = q_{\rho_1,\rho_2}$ and obtain

$$\sum_{|u_{\rho_2}| \leq k_2 M_2^{\rho_2}} \min \Big\{ 2d_1 M_1^{\rho_1}, \frac{1}{\|\beta^1_{\rho_1}(u)\|} \Big\} \lesssim M_1^{\rho_1} + q \log q + \frac{M_1^{\rho_1} M_2^{\rho_2}}{q} + M_2^{\rho_2} \log q$$

$$\lesssim M_1^{\rho_1} M_2^{\rho_2} \Big(\frac{1}{M_2^{\rho_2}} + \frac{q \log q}{M_1^{\rho_1} M_2^{\rho_2}} + \frac{1}{q} + \frac{\log q}{M_1^{\rho_1}} \Big).$$

Hence

$$\sum_{u \in \mathcal{B}_{d_2}(M_2)} \Big| \sum_{v \in \mathcal{B}_{d_1}(M_1)} \boldsymbol{e}(\beta^1(u) \cdot v) \Big| \lesssim \Big(\prod_{j=1}^{2} M_j^{\frac{d_j(d_j+1)}{2}} \Big) \Big(\frac{1}{M_2^{\rho_2}} + \frac{q \log q}{M_1^{\rho_1} M_2^{\rho_2}} + \frac{1}{q} + \frac{\log q}{M_1^{\rho_1}} \Big),$$

establishing (5.31).

Step 5. We use the bound (5.31) in (5.30) to conclude

$$S^1_{K_1,M_1,K_2,M_2}(Q)^{4k_1 k_2} \lesssim M_1^{4k_1 k_2 - 2k_1} M_2^{4k_1 k_2 - 2k_2} J_{k_1,d_1}(M_1) J_{k_2,d_2}(M_2)$$

$$\times M_1^{\frac{d_1(d_1+1)}{2}} M_2^{\frac{d_2(d_2+1)}{2}} \Big(\frac{1}{M_2^{\rho_2}} + \frac{q \log q}{M_1^{\rho_1} M_2^{\rho_2}} + \frac{1}{q} + \frac{\log q}{M_1^{\rho_1}} \Big).$$

From Vinogradov's mean value theorem (or more precisely from (5.14) with $s = k_i := d_i(d_i + 1)$ and $k = d_i$ for $i \in [2]$), we conclude from (5.14), $J_{k_i,d_i}(M_i) \leq C M_i^{3k_i/2}, i = 1, 2$ and so

$$S^1_{K_1,M_1,M_2,M_2}(Q)^{4k_1 k_2} \lesssim M_1^{4k_1 k_2} M_2^{4k_1 k_2} \Big(\frac{1}{M_2^{\rho_2}} + \frac{q \log q}{M_1^{\rho_1} M_2^{\rho_2}} + \frac{1}{q} + \frac{\log q}{M_1^{\rho_1}} \Big).$$

In a similar way, using (5.32) in (5.29), we also have

$$S^1_{K_1,M_1,M_2,M_2}(Q)^{4k_1 k_2} \lesssim M_1^{4k_1 k_2} M_2^{4k_1 k_2} \Big(\frac{1}{M_1^{\rho_1}} + \frac{q \log q}{M_1^{\rho_1} M_2^{\rho_2}} + \frac{1}{q} + \frac{\log q}{M_2^{\rho_2}} \Big).$$

Therefore, $S^1_{K_1,M_1,M_2,M_2}(Q)^{4k_1 k_2}$ is bounded from above by the minimum of these two bounds. By Remark 5.25, this completes the proof of Proposition 5.22. □

5.2. Double Weyl's inequality in the Newton diagram sectors

Throughout this subsection, we assume that $P \in \mathbb{Z}[m_1, m_2]$ and $P(0,0) = 0$. Moreover, we assume that P is non-degenerate in the sense of (1.14); see the remark below Theorem 1.11. Then for every $\xi \in \mathbb{R}$, we define a corresponding polynomial $P_\xi \in \mathbb{R}[m_1, m_2]$ by setting

$$P_\xi(m_1, m_2) := \xi P(m_1, m_2). \tag{5.33}$$

It is clear to see that the backwards Newton diagrams of P and P_ξ are the same $N_P = N_{P_\xi}$. Let $r \in \mathbb{Z}_+$ be the number of vertices in the backwards Newton diagram N_P. In view of (4.7) and (4.8) from Remark 4.5 for $r \geq 2$, we have

$$
\begin{aligned}
\log M_1 &\lesssim \log M_2 && \text{if} && (M_1, M_2) \in \mathbb{S}_\tau(1), \\
\log M_1 &\simeq \log M_2 && \text{if} && (M_1, M_2) \in \mathbb{S}_\tau(j) \text{ for } 1 < j < r, \\
\log M_2 &\lesssim \log M_1 && \text{if} && (M_1, M_2) \in \mathbb{S}_\tau(r).
\end{aligned}
\tag{5.34}
$$

Consequently, we may define a quantity $M^*_{r,j}$ as follows. If $r = 1$, we simply set

$$M^*_{1,1} := M_1 \vee M_2 \quad \text{if} \quad (M_1, M_2) \in \mathbb{S}_\tau(1) = \mathbb{D}_\tau \times \mathbb{D}_\tau. \tag{5.35}$$

If $r \geq 2$, we set

$$M^*_{r,j} := \begin{cases} M_2 & \text{if } (M_1, M_2) \in \mathbb{S}_\tau(1) \text{ for } j = 1, \\ M_1 \vee M_2 & \text{if } (M_1, M_2) \in \mathbb{S}_\tau(j) \text{ for } 1 < j < r, \\ M_1 & \text{if } (M_1, M_2) \in \mathbb{S}_\tau(r) \text{ for } j = r. \end{cases} \tag{5.36}$$

The quantity $\log M^*_{r,j}$ will always allow us to extract the larger parameter (larger up to a multiplicative constant as in (5.34)) from $\log M_1$ and $\log M_2$. We estimate $|S_{K_1, M_1, K_2, M_2}(P_\xi)|$ in terms of $\log M^*_{r,j}$, whenever $(M_1, M_2) \in \mathbb{S}_\tau(j)$ for $j \in [r]$, and $(K_1, K_2) \in \mathbb{N}^2$ satisfying $M_1 \lesssim K_1 \leq M_1$ and $M_2 \lesssim K_2 \leq M_2$.

Proposition 5.37. *Let $P_\xi \in \mathbb{R}[m_1, m_2]$ be the polynomial in (5.33) corresponding to a polynomial $P \in \mathbb{Z}[m_1, m_2]$ with the properties above. Let $r \in \mathbb{Z}_+$ be the number of vertices in the backwards Newton diagram N_P. Let $\tau > 1$, $\alpha > 1$, $j \in [r]$ be given. Let $v_j = (v_{j,1}, v_{j,2})$ be the vertex of the backwards Newton diagram N_P. Then there exists a constant $\beta_0 := \beta_0(\alpha) > \alpha$ such that for every $\beta \in (\beta_0, \infty) \cap \mathbb{Z}_+$, we find a constant $0 < C = C(\alpha, \beta_0, \beta, j, \tau, P) < \infty$ such that for every $(M_1, M_2) \in \mathbb{S}_\tau(j)$ and $(K_1, K_2) \in \mathbb{N}^2$ satisfying $M_1 \lesssim K_1 \leq M_1$ and $M_2 \lesssim K_2 \leq M_2$ the following holds. Suppose that there are $a \in \mathbb{Z}, q \in \mathbb{Z}_+$ such that $(a, q) = 1$ and*

$$(\log M^*_{r,j})^\beta \lesssim q \leq M_1^{v_{j,1}} M_2^{v_{j,2}} (\log M^*_{r,j})^{-\beta}, \tag{5.38}$$

and

$$\left| \xi - \frac{a}{q} \right| \leq \frac{(\log M^*_{r,j})^\beta}{q M_1^{v_{j,1}} M_2^{v_{j,2}}}, \tag{5.39}$$

where $M^*_{r,j}$ is defined in (5.36). Then one has

$$|S_{K_1, M_1, K_2, M_2}(P_\xi)| \leq C M_1 M_2 (\log M^*_{r,j})^{-\alpha}. \tag{5.40}$$

Proof. We note that the following three scenarios may occur when $r > 1$:

1. If $j = 1$, we have $v_{1,1} = 0$ or $v_1 \in \mathbb{Z}_+ \times \mathbb{Z}_+$. In this case, we also have $\log M_1 \lesssim \log M_2$.
2. If $j = r$, we have $v_{r,2} = 0$ or $v_r \in \mathbb{Z}_+ \times \mathbb{Z}_+$. In this case, we also have $\log M_1 \gtrsim \log M_2$.
3. If $1 < j < r$, we have $v_j \in \mathbb{Z}_+ \times \mathbb{Z}_+$. In this case, we also have $\log M_1 \simeq \log M_2$.

Note that if $r = 1$, then $\mathbb{S}_\tau(1) = \mathbb{D}_\tau \times \mathbb{D}_\tau$ and $v_1 \in \mathbb{Z}_+ \times \mathbb{Z}_+$, since P is non-degenerate in the sense of (1.14). Throughout the proof, in the case of $r = 1$, we will additionally assume that $\log M_1 \leq \log M_2$. Taking into account (5.35) and (5.36), we can also assume that $\log M_1 \vee \log M_2$ is sufficiently large (i.e., $\log M_1 \vee \log M_2 > C_0$, where $C_0 = C_0(\alpha, \beta_0, j, \tau, P) > 0$ is a large absolute constant). Otherwise, inequality (5.40) follows. The proof will be divided into three steps.

Step 1. We first establish (5.40) when $j = 1$ and $v_{1,1} = 0$ or $j = r$ and $v_{r,2} = 0$. Suppose that $j = 1$ and $v_{1,1} = 0$ holds. The case when $j = r$ and $v_{r,2} = 0$ can be proved in a similar way, so we omit the details. As we have seen above, $\log M_1 \lesssim \log M_2$. By (5.38) and (5.39), we obtain

$$\left| \xi - \frac{a}{q} \right| \leq \frac{1}{q^2}.$$

Applying Lemma 5.3 with $Q = c_{0,v_{1,2}}$ and $M = q$, we may find a fraction a'/q' such that $(a', q') = 1$ and $q/(2c_{0,v_{1,2}}) \leq q' \leq 2q$ and

$$\left| c_{0,v_{1,2}} \xi - \frac{a'}{q'} \right| \leq \frac{1}{2q'q} \leq \frac{1}{(q')^2}.$$

Thus, by Proposition 5.4, noting that $v_{1,2} \geq 1$, we obtain

$$|S_{K_1,M_1,K_2,M_2}(P_\xi)| \leq S^1_{K_1,M_1,K_2,M_2}(P_\xi)$$
$$\lesssim M_1 M_2 \log(M_2) \left(\frac{1}{q'} + \frac{1}{M_2} + \frac{q'}{M_2^{v_{1,2}}} \right)^{\frac{1}{\tau(\deg P)}}$$
$$\lesssim M_1 M_2 (\log M^*_{r,j})^{-\frac{\beta}{\tau(\deg P)}+1},$$

since $\log M^*_{r,j} \simeq \log M_2$. It suffices to take $\beta > \tau(\deg P)(\alpha + 1)$ and the claim in (5.40) follows.

Step 2. We now establish (5.40) when $1 \leq j \leq r$ and $v_j \in \mathbb{Z}_+ \times \mathbb{Z}_+$ (note that when $1 < j < r$, we automatically have $v_j \in \mathbb{Z}_+ \times \mathbb{Z}_+$). If $r = 1$, then we assume that $\log M_1 \leq \log M_2$. If $r \geq 2$, we will assume that $1 \leq j < r$, which gives that $\log M_1 \lesssim \log M_2$. The case when $j = r$ can be proved in much the same way (with the difference that $\log M_1 \gtrsim \log M_2$), we omit the details. In this step, we additionally assume that $M_1 \leq (\log M^*_{r,j})^\chi$ for some $0 < \chi < \beta/(8 \deg P)$ with β to be specified later.

Notice that (5.38) and (5.39) imply

$$\left| \xi - \frac{a}{q} \right| \leq \frac{1}{q^2}.$$

By (5.38) and $M_1 \leq (\log M^*_{r,j})^\chi$, we conclude

$$(\log M^*_{r,j})^\beta \leq q \leq M_2^{v_{j,2}} (\log M^*_{r,j})^{-3\beta/4}$$

since $\chi < \beta/(8 \deg P)$. We note that the polynomial P can be written as

$$P(m_1, m_2) = P_{v_{j,1}}(m_1) m_2^{v_{j,2}} + \sum_{\substack{(\gamma_1, \gamma_2) \in S_P \\ \gamma_2 \neq v_{j,2}}} c_{\gamma_1, \gamma_2} m_1^{\gamma_1} m_2^{\gamma_2},$$

where $P_{v_{j,1}} \in \mathbb{Z}[m_1]$ and $\deg P_{v_{j,1}} = v_{j,1}$.

Observe that for every $1 \leq m_1 \leq M_1 \leq (\log M^*_{r,j})^\chi$, one has

$$|P_{v_{j,1}}(m_1)| \leq \#S_P \max_{(\gamma_1, \gamma_2) \in S_P} |c_{\gamma_1, \gamma_2}| M_1^{\deg P} \lesssim_P (\log M^*_{r,j})^{\beta/4}.$$

Applying Lemma 5.3 with $M = M_2^{v_{j,2}}(\log M_{r,j}^*)^{-3\beta/4}$ and $Q = P_{v_{j,1}}(m_1)$ for each $K_1 < m_1 \leq M_1$ (noting that $P_{v_{j,1}}(m_1) \neq 0$ for large m_1), we find a fraction a'/q' so that $(a', q') = 1$ and $(\log M_{r,j}^*)^{3\beta/4} \lesssim q' \leq 2M_2^{v_{j,2}}(\log M_{r,j}^*)^{-3\beta/4}$ and

$$\left| P_{v_{j,1}}(m_1)\xi - \frac{a'}{q'} \right| \leq \frac{(\log M_{r,j}^*)^{3\beta/4}}{2q' M_2^{v_{j,2}}} \leq \frac{1}{(q')^2}.$$

We apply Proposition 5.4 for each $1 \leq m_1 \leq M_1$, noting that $v_{j,2} \geq 1$, to bound

$$\left| \sum_{m_2 = K_2+1}^{M_2} e(P_\xi(m_1, m_2)) \right| \lesssim M_2 \log(M_2) \left(\frac{1}{q'} + \frac{1}{M_2} + \frac{q'}{M_2^{v_{j,2}}} \right)^{\frac{1}{\tau(\deg P)}} \lesssim M_2 (\log M_{r,j}^*)^{-\frac{3\beta}{4\tau(\deg P)}+1},$$

since $\log M_{r,j}^* \simeq \log M_2$ for $j \in [r-1]$. It suffices to take $\beta > \frac{4}{3}\tau(\deg P)(\alpha + 1)$ and (5.40) follows.

Step 3. As in the previous step, $1 \leq j < r$ (or $r = 1$ and $\log M_1 \leq \log M_2$) and we now assume that $(\log M_{r,j}^*)^\chi \leq M_1 \lesssim M_2$ for some $0 < \chi < \beta/(8 \deg P)$, which will be further adjusted. The case when $j = r$ can be established in a similar fashion keping in mind that $\log M_1 \gtrsim \log M_2$. In fact, we take $\chi := \beta/(16 \deg P) + 1$, which forces $\beta > 16 \deg P$.

Applying Lemma 5.3 with $Q = c_{v_{j,1}, v_{j,2}}$ and $M = q$, we find a fraction a'/q' so that $(a', q') = 1$ and $(\log M_{r,j}^*)^\beta \lesssim_P q(2Q)^{-1} \leq q' \leq 2q$ and

$$\left| c_{v_{j,1}, v_{j,2}}\xi - \frac{a'}{q'} \right| \leq \frac{1}{(q')^2}.$$

From Proposition 5.22, we obtain (with $M_- = \min(M_1^{v_{j,1}}, M_2^{v_{j,2}})$ and $M_+ = \max(M_1^{v_{j,1}}, M_2^{v_{j,2}})$)

$$|S_{K_1, M_1, K_2, M_2}(P_\xi)| \lesssim M_1 M_2 \left(\frac{1}{M_-} + \frac{q' \log q'}{M_1^{v_{j,1}} M_2^{v_{j,2}}} + \frac{1}{q'} + \frac{\log q'}{M_+} \right)^{\frac{1}{4(1+\deg P)^5}}$$

$$\lesssim M_1 M_2 (\log M_{r,j}^*)^{-\frac{\beta}{64(1+\deg P)^5}}.$$

Taking $\beta > 64(1 + \deg P)^5(\alpha + 1)$, we obtain (5.40). This completes the proof of Proposition 5.37. □

5.3. Estimates for double complete exponential sums

In this subsection, we provide estimates for double complete exponential sums in the spirit of Gauss. We begin with a well-known bound which is also a simple consequence of Proposition 5.22.

Lemma 5.41 [1]. *Let $P \in \mathbb{Q}[m_1, m_2]$ be a polynomial as in (4.1) and let $a_{\gamma_1, \gamma_2} \in \mathbb{Z}$ and $q \in \mathbb{Z}_+$ satisfy $c_{\gamma_1, \gamma_2} = a_{\gamma_1, \gamma_2}/q$ for each $(\gamma_1, \gamma_2) \in S_P$ such that*

$$\gcd(\{a_{\gamma_1, \gamma_2} : (\gamma_1, \gamma_2) \in S_P\} \cup \{q\}) = 1.$$

Consider the exponential sum $S_{q,q}$ from (5.18). Then there are $C > 0$ and $\delta \in (0, 1)$ such that

$$|S_{q,q}(P)| \leq Cq^{2-\delta} \tag{5.42}$$

holds. The constant C can be taken to depend only on the degree of P.

We now derive simple consequences of Lemma 5.41 for exponential sums that arise in the proof of our main result. Let $P \in \mathbb{Z}[m_1, m_2]$ be such that

$$P(m_1, m_2) := \sum_{(\gamma_1, \gamma_2) \in S_P} c^P_{\gamma_1, \gamma_2} m_1^{\gamma_1} m_2^{\gamma_2}, \tag{5.43}$$

where $c^P_{(0,0)} = 0$. We additionally assume that P is non-degenerate (see the remark below Theorem 1.11). That is, we have $S_P \cap (\mathbb{Z}_+ \times \mathbb{Z}_+) \neq \emptyset$. Using the definition of P_ξ from (5.33), we define the complete exponential sum by

$$G(a/q) := \frac{1}{q^2} \sum_{r_1=1}^{q} \sum_{r_2=1}^{q} e(P_{a/q}(r_1, r_2)), \qquad a/q \in \mathbb{Q}, \tag{5.44}$$

and we also have partial complete exponential sums defined by

$$G^1_{m_1}(a/q) := \frac{1}{q} \sum_{r_2=1}^{q} e(P_{a/q}(m_1, r_2)), \qquad a/q \in \mathbb{Q}, \ m_1 \in \mathbb{Z},$$

$$G^2_{m_2}(a/q) := \frac{1}{q} \sum_{r_1=1}^{q} e(P_{a/q}(r_1, m_2)), \qquad a/q \in \mathbb{Q}, \ m_2 \in \mathbb{Z}. \tag{5.45}$$

Proposition 5.46. *Let $P \in \mathbb{Z}[m_1, m_2]$ be a polynomial as in (5.43) which is non-degenerate (that is, $S_P \cap (\mathbb{Z}_+ \times \mathbb{Z}_+) \neq \emptyset$). Then there is $C_P > 0$ and $\delta \in (0, 1)$ such that the following inequalities hold. If $a/q \in \mathbb{Q}$ and $(a, q) = 1$, then*

$$|G(a/q)| \leq C_P \, q^{-\delta}. \tag{5.47}$$

Moreover, for every sufficiently large $K_1, M_1 \in \mathbb{Z}_+$ depending on P, one has

$$\frac{1}{M_1} \sum_{m_1 = K_1+1}^{M_1} |G^1_{m_1}(a/q)| \leq C_P \, q^{-\delta}, \tag{5.48}$$

and similarly, for every sufficiently large $K_2, M_2 \in \mathbb{Z}_+$ depending on P, one has

$$\frac{1}{M_2} \sum_{m_2 = K_2+1}^{M_2} |G^2_{m_2}(a/q)| \leq C_P \, q^{-\delta}. \tag{5.49}$$

Proof. We prove Proposition 5.46 in two steps.

Step 1. In this step, we establish (5.47). Fix $a/q \in \mathbb{Q}$ such that $(a, q) = 1$. For any $(\gamma_1, \gamma_2) \in S_P$, we let $a_{\gamma_1, \gamma_2} := ac^P_{\gamma_1, \gamma_2}/(c^P_{\gamma_1, \gamma_2}, q)$ and $q_{\gamma_1, \gamma_2} := q/(c^P_{\gamma_1, \gamma_2}, q)$. Now with this notation, we see that

$$P_{a/q}(r_1, r_2) = Q(r_1, r_2) := \sum_{\gamma_1=0}^{d_1} \sum_{\gamma_2=0}^{d_2} \frac{a_{\gamma_1, \gamma_2}}{q_{\gamma_1, \gamma_2}} r_1^{\gamma_1} r_2^{\gamma_2},$$

for some integers $d_1, d_2 \geq 1$. Furthermore, $G(a/q) = q^{-2} S_{q,q}(Q)$; see (5.44). We take $(\rho_1, \rho_2) \in S_P \cap (\mathbb{Z}_+ \times \mathbb{Z}_+) \neq \emptyset$ and use (5.42), which yields

$$|G(a/q)| = q^{-2} |S_{q,q}(Q)| \lesssim_P q^{-\delta}.$$

This completes the proof of (5.47).

Step 2. We only prove (5.48); the proof of (5.49) is exactly the same. We fix $a/q \in \mathbb{Q}$ such that $(a, q) = 1$, and we also fix $(\rho_1, \rho_2) \in S_P \cap (\mathbb{Z}_+ \times \mathbb{Z}_+) \neq \emptyset$. Using Lemma 5.3, we find a reduced fraction $a_{\rho_1,\rho_2}/q_{\rho_1,\rho_2}$ so that $(a_{\rho_1,\rho_2}, q_{\rho_1,\rho_2}) = 1$ and

$$\left| \frac{ac_{\rho_1,\rho_2}^P}{q} - \frac{a_{\rho_1,\rho_2}}{q_{\rho_1,\rho_2}} \right| \leq \frac{1}{2q_{\rho_1,\rho_2}q}$$

with $q/(2c_{\rho_1,\rho_2}) \leq q_{\rho_1,\rho_2} \leq 2q$. We fix $\chi > 0$ and assume first that $M_1 \geq q^\chi$. Appealing to inequality (5.24) with $M_2 = q$, we obtain for some $\delta \in (0, 1)$ that

$$\frac{1}{M_1} \sum_{m_1=K_1+1}^{M_1} |G_{m_1}^1(a/q)| \lesssim_P q^{-\delta}.$$

We now establish a similar bound assuming that $M_1 < q^\chi$ for a sufficiently small $\chi > 0$, which will be specified momentarily. Our polynomial P from (5.43) can be rewritten as

$$P(m_1, m_2) = \sum_{\gamma_2=1}^{d_2} P_{\gamma_2}(m_1) m_2^{\gamma_2} + P_0(m_1),$$

for some $d_2 \geq 1$ where $P_{\gamma_2} \in \mathbb{Z}[m_1]$ and $\deg P_{\gamma_2} \leq \deg P$. Take $0 < \chi < \frac{1}{10 \deg P}$ and observe that for every $1 \leq \gamma_2 \leq d_2$ and for every $1 \leq m_1 \leq M_1 \leq q^\chi$, one has

$$|P_{\gamma_2}(m_1)| \leq \#S_P \max_{(\gamma_1,\gamma_2) \in S_P} |c_{\gamma_1,\gamma_2}| M_1^{\deg P} \leq q^{1/4}, \tag{5.50}$$

whenever q is sufficiently large in terms of the coefficients of P.

Assume first that $d_2 \geq 2$, and we may take $\rho_2 = d_2$. Applying Lemma 5.3 with $Q = P_{\rho_2}(m_1)$ for each $K_1 < m_1 \leq M_1$ (noting that $P_{\rho_2}(m_1) \neq 0$ for sufficiently large $m_1 \geq K_1$), we find a fraction a'/q' so that $(a', q') = 1$ and $\frac{1}{2}q^{3/4} \leq q' \leq 2q$ and

$$\left| P_{\rho_2}(m_1) \frac{a}{q} - \frac{a'}{q'} \right| \leq \frac{1}{2q'q} \leq \frac{1}{(q')^2}.$$

Then we apply Proposition 5.4 for each $K_1 < m_1 \leq M_1$, which gives

$$|G_{m_1}^1(a/q)| \lesssim \log(2q) \left(\frac{1}{q'} + \frac{1}{q} + \frac{q'}{q^{d_2}} \right)^{\frac{1}{\tau(d_2)}} \lesssim (\log q) q^{-\frac{3}{4\tau(d_2)}} \lesssim q^{-\delta},$$

for some $\delta \in (0, 1)$ and (5.48) follows, since $d_2 \geq 2$.

Assume now that $d_2 = 1$. Then

$$\frac{1}{M_1} \sum_{m_1=K_1+1}^{M_1} |G_{m_1}^1(a/q)| = \frac{1}{M_1} \#\{K_1 < m_1 \leq M_1 : P_1(m_1) \equiv 0 \bmod q\} = 0,$$

in view of (5.50), which ensures that $\{m_1 \in [M_1] : P_1(m_1) \equiv 0 \bmod q\} = \emptyset$. $\qquad\square$

6. Multi-parameter Ionescu–Wainger theory

One of the most important ingredients in our argument is the Ionescu–Wainger multiplier theorem [34] (see also [46]), and its vector-valued variant from [52] (see also [62]). We begin with recalling the results from [34] and [52] and fixing necessary notation and terminology.

6.1. Ionescu–Wainger multiplier theorem

Let $\mathbb{P}$ be the set of all prime numbers, and let $\rho \in (0, 1)$ be a sufficiently small absolute constant. We then define the natural number

$$D := D_\rho := \lfloor 2/\rho \rfloor + 1,$$

and for any integer $l \in \mathbb{N}$, set

$$N_0 := N_0^{(l)} := \lfloor 2^{\rho l/2} \rfloor + 1, \quad \text{and} \quad Q_0 := Q_0^{(l)} := (N_0!)^D.$$

We also define the set

$$P_{\leq l} := \big\{ q = Qw : Q | Q_0 \text{ and } w \in W_{\leq l} \cup \{1\} \big\},$$

where

$$W_{\leq l} := \bigcup_{k \in [D]} \bigcup_{(\gamma_1, \ldots, \gamma_k) \in [D]^k} \big\{ p_1^{\gamma_1} \cdots p_k^{\gamma_k} : p_1, \ldots, p_k \in (N_0^{(l)}, 2^l] \cap \mathbb{P} \text{ are distinct} \big\}.$$

In other words, $W_{\leq l}$ is the set of all products of prime factors from $(N_0^{(l)}, 2^l] \cap \mathbb{P}$ of length at most D, at powers between 1 and D.

Remark 6.1. For every $\rho \in (0, 1)$, there exists a large absolute constant $C_\rho \geq 1$ such that the following elementary facts about the sets $P_{\leq l}$ hold:

(i) If $l_1 \leq l_2$, then $P_{\leq l_1} \subseteq P_{\leq l_2}$.
(ii) One has $[2^l] \subseteq P_{\leq l} \subseteq [2^{C_\rho 2^{\rho l}}]$.
(iii) If $q \in P_{\leq l}$, then all factors of q also lie in $P_{\leq l}$.
(iv) One has $Q_{\leq l} := \text{lcm}(P_{\leq l}) \lesssim 2^{C_\rho 2^l}$.

By property (i), it makes sense to define $P_l := P_{\leq l} \backslash P_{\leq l-1}$, with the convention that $P_{\leq l}$ is empty for negative l. From property (ii), for all $q \in P_l$, we have

$$2^{l-1} < q \leq 2^{C_\rho 2^{\rho l}}. \tag{6.2}$$

Let $d \in \mathbb{Z}_+$ and define 1-periodic sets

$$\Sigma_{\leq l}^d := \left\{ \frac{a}{q} \in (\mathbb{Q} \cap \mathbb{T})^d : q \in P_{\leq l} \text{ and } (a, q) = 1 \right\}, \quad \text{and} \quad \Sigma_l^d := \Sigma_{\leq l}^d \backslash \Sigma_{\leq l-1}^d, \tag{6.3}$$

where $(a, q) = (a_1, \ldots, a_d, q) = 1$ for any $a = (a_1, \ldots, a_d) \in \mathbb{Z}^d$. Then by (6.2), we see

$$\#\Sigma_{\leq l}^d \leq 2^{C_\rho(d+1)2^{\rho l}}. \tag{6.4}$$

Let $k \in \mathbb{Z}_+$ be fixed. For any finite family of fractions $\Sigma \subseteq (\mathbb{T} \cap \mathbb{Q})^k$ and a measurable function $\mathfrak{m} : \mathbb{R}^k \to B$ taking its values in a separable Banach space B which is supported on the unit cube $[-1/2, 1/2]^k$, define a 1-periodic extension of $\mathfrak{m}$ by

$$\Theta_\Sigma[\mathfrak{m}](\xi) := \sum_{a/q \in \Sigma} \mathfrak{m}(\xi - a/q), \quad \xi \in \mathbb{T}^k.$$

We will also need to introduce the notion of Γ-lifted extensions of $\mathfrak{m}$. For $d \in \mathbb{Z}_+$ consider $\Gamma := \{i_1, \ldots, i_k\} \subseteq [d]$ of size $k \in [d]$. We define a Γ-lifted 1-periodic extension of $\mathfrak{m}$ by

$$\Theta_{\Sigma}^{\Gamma}[\mathfrak{m}](\xi) := \sum_{a/q \in \Sigma} \mathfrak{m}(\xi_{i_1} - a_1/q, \ldots, \xi_{i_k} - a_k/q), \quad \text{for} \quad \xi = (\xi_1, \ldots, \xi_d) \in \mathbb{T}^d.$$

We now recall the following vector-valued Ionescu–Wainger multiplier theorem from [52, 62].

Theorem 6.5. *Let $d \in \mathbb{Z}_+$ be given. For every $\rho \in (0,1)$ and for every $p \in (1, \infty)$, there exists an absolute constant $C_{p,\rho,d} > 0$, that depends only on p, ρ and d, such that, for every $l \in \mathbb{N}$, the following holds. Let $0 < \varepsilon_l \le 2^{-10 C_\rho 2^{\rho l}}$, and let $\mathfrak{m} : \mathbb{R}^d \to L(H_0, H_1)$ be a measurable function supported on $\varepsilon_l[-1/2, 1/2]^d$, with values in the space $L(H_0, H_1)$ of bounded linear operators between separable Hilbert spaces H_0 and H_1. Let*

$$\mathbf{A}_p := \|T_{\mathbb{R}^d}[\mathfrak{m}]\|_{L^p(\mathbb{R}^d; H_0) \to L^p(\mathbb{R}^d; H_1)}. \tag{6.6}$$

Then the 1-periodic multiplier

$$\Theta_{\Sigma_{\le l}^d}[\mathfrak{m}](\xi) = \sum_{a/q \in \Sigma_{\le l}^d} \mathfrak{m}(\xi - a/q) \quad \text{for} \quad \xi \in \mathbb{T}^d, \tag{6.7}$$

where $\Sigma_{\le l}^d$ is the set of all reduced fractions in (6.3), satisfies

$$\|T_{\mathbb{Z}^d}[\Theta_{\Sigma_{\le l}^d}[\mathfrak{m}]] f\|_{\ell^p(\mathbb{Z}^d; H_1)} \le C_{p,\rho,d} \mathbf{A}_p \|f\|_{\ell^p(\mathbb{Z}^d; H_0)} \tag{6.8}$$

for every $f \in \ell^p(\mathbb{Z}^d; H_0)$.

The advantage of applying Theorem 6.5 is that one can directly transfer square function estimates from the continuous to the discrete setting, which will be useful in Section 7. The hypothesis (6.6), unlike the support hypothesis, is scale-invariant, in the sense that the constant $\mathbf{A}_p$ does not change when $\mathfrak{m}$ is replaced by $\mathfrak{m}(A\cdot)$ for any invertible linear transformation $A : \mathbb{R}^d \to \mathbb{R}^d$.

Theorem 6.5 was originally established by Ionescu and Wainger [34] in the scalar-valued setting with an extra factor $(l + 1)^D$ in the right-hand side of (6.8). Their proof is based on an intricate inductive argument that exploits super-orthogonality phenomena. A slightly different proof with factor $(l + 1)$ in (6.8) was given in [46]. The latter proof, instead of induction as in [34], used certain recursive arguments, which clarified the role of the underlying square functions and orthogonalities (see also [52, Section 2]). The theorem in the context of super-orthogonality phenomena is discussed in a survey by Pierce [55] in a much broader context. Finally, we refer to the recent paper of Tao [62], where Theorem 6.5 as stated above, with a uniform constant $\mathbf{A}_p$, is established.

For future reference, we also recall the sampling principle of Magyar–Stein–Wainger from [44], which was an important ingredient in the proof of Theorem 6.5.

Proposition 6.9. *Let $d \in \mathbb{Z}_+$ be given. There exists an absolute constant $C > 0$ such that the following holds. Let $p \in [1, \infty]$ and $q \in \mathbb{Z}_+$, and let B_1, B_2 be finite-dimensional Banach spaces. Let $\mathfrak{m} : \mathbb{R}^d \to L(B_1, B_2)$ be a bounded operator-valued function supported on $[-1/2, 1/2]^d/q$ and let $\mathfrak{m}_{\mathrm{per}}^q$ be the periodic multiplier*

$$\mathfrak{m}_{\mathrm{per}}^q(\xi) := \sum_{n \in \mathbb{Z}^d} \mathfrak{m}(\xi - n/q), \qquad \xi \in \mathbb{T}^d.$$

Then

$$\|T_{\mathbb{Z}^d}[\mathfrak{m}_{\mathrm{per}}^q]\|_{\ell^p(\mathbb{Z}^d; B_1) \to \ell^p(\mathbb{Z}^d; B_2)} \le C \|T_{\mathbb{R}^d}[\mathfrak{m}]\|_{L^p(\mathbb{R}^d; B_1) \to L^p(\mathbb{R}^d; B_2)}.$$

The proof can be found in [44, Corollary 2.1, pp. 196]. We also refer to [50] for a generalization of Proposition 6.9 to real interpolation spaces. We emphasize that B_1 and B_2 are general (finite dimensional) Banach spaces in Proposition 6.9, in contrast to the Hilbert space-valued multipliers appearing in Theorem 6.5, and so Proposition 6.9 includes maximal function formulations and can also accommodate oscillation semi-norms.

6.2. One-parameter semi-norm variant of Theorem 6.5

Let $\Lambda := \{\lambda_1, \ldots, \lambda_k\} \subset \mathbb{Z}_+$ be a set of size $k \in \mathbb{Z}_+$ of natural exponents, and consider the associated one-parameter family of dilations which for every $x \in \mathbb{R}^k$, is defined by

$$(0, \infty) \ni t \mapsto t \circ x := (t^{\lambda_1} x_1, \ldots, t^{\lambda_k} x_k) \in \mathbb{R}^k.$$

Let $\Upsilon := (\Upsilon_n : \mathbb{R}^k \to \mathbb{C} : n \in \mathbb{N})$ be a sequence of measurable functions which define a positive sequence of operators in the sense that for every $n \in \mathbb{N}$, one has

$$T_{\mathbb{R}^k}[\Upsilon_n] f \ge 0 \quad \text{if} \quad f \ge 0. \tag{6.10}$$

Furthermore, suppose there exist $C_\Upsilon > 0$, $0 < \delta_\Upsilon < 1$ and $1 < \tau \le 2$ such that for every $\xi \in \mathbb{R}^k$ and $n \in \mathbb{N}$, one has

$$|\Upsilon_n(\xi)| \le C_\Upsilon \min\left\{1, |\tau^n \circ \xi|^{-\delta_\Upsilon}\right\}, \tag{6.11}$$

$$|\Upsilon_n(\xi) - 1| \le C_\Upsilon \min\left\{1, |\tau^n \circ \xi|^{\delta_\Upsilon}\right\}. \tag{6.12}$$

The condition (6.10) implies that the operator $T_{\mathbb{R}^k}[\Upsilon_n] f = f * \mu_n$ is convolution with positive measure μ_n and condition (6.12) implies $\Upsilon_n(0) = 1$ and so each μ_n is a probability measure. Hence, for every $p \in [1, \infty)$,

$$A_p^\Upsilon := \sup_{n \in \mathbb{N}} \|T_{\mathbb{R}^k}[\Upsilon_n]\|_{L^p(\mathbb{R}^k) \to L^p(\mathbb{R}^k)} \le 1. \tag{6.13}$$

In this generality, $L^p(\mathbb{R}^k)$ estimates with $1 < p \le \infty$ for the maximal function $\sup_{n \in \mathbb{N}} |T_{\mathbb{R}^k}[\Upsilon_n] f(x)|$ were obtained in [22] and corresponding r-variational and jump inequalites were established in [38] (see also [51]). Here we extend these results further.

For $d \in \mathbb{Z}_+$, consider $\Gamma := \{i_1, \ldots, i_k\} \subseteq [d]$ of size $k \in [d]$ and define a Γ-lifted sequence of measurable functions $\Upsilon^\Gamma := (\Upsilon_n^\Gamma : \mathbb{R}^d \to \mathbb{C} : n \in \mathbb{N})$ by setting

$$\Upsilon_n^\Gamma(\xi) := \Upsilon_n(\xi_{i_1}, \ldots, \xi_{i_k}) \quad \text{for} \quad \xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d.$$

Our first main result is the following one-parameter semi-norm variant of Theorem 6.5.

Theorem 6.14. *Let $d \in \mathbb{Z}_+$ and $\Gamma \subseteq [d]$ of size $k \in [d]$ be given. Let $\Upsilon = (\Upsilon_n : \mathbb{R}^k \to \mathbb{C} : n \in \mathbb{N})$ be a sequence of measurable functions satisfying conditions (6.10), (6.11) and (6.12), and let $\Upsilon^\Gamma := (\Upsilon_n^\Gamma : \mathbb{R}^d \to \mathbb{C} : n \in \mathbb{N})$ be the corresponding Γ-lifted sequence. For every $\rho \in (0, 1)$ and for every $p \in (1, \infty)$, there exists an absolute constant $0 < C = C(d, p, \rho, \tau, \Gamma, A_p^\Upsilon, C_\Upsilon) < \infty$ such that for every integer $l \in \mathbb{N}$ and $m \le -10 C_\rho 2^{2\rho l}$, the following holds. If*

$$\operatorname{supp} \Upsilon_n \subseteq 2^m [-1/2, 1/2]^k \quad \text{for all} \quad n \in \mathbb{N}, \tag{6.15}$$

then for every $f = (f_\iota : \iota \in \mathbb{N}) \in \ell^p(\mathbb{Z}^d; \ell^2(\mathbb{N}))$, one has

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{N})} \left\| \left(\sum_{\iota \in \mathbb{N}} O_{I,J}\big(T_{\mathbb{Z}^d}\big[\Theta_{\Sigma^d_{\leq l}}[\Upsilon^\Gamma_n \eta^{\Gamma^c}_{\leq m}]\big] f_\iota : n \in \mathbb{N}\big)^2 \right)^{1/2} \right\|_{\ell^p(\mathbb{Z}^d)} \leq C(l+1)\|f\|_{\ell^p(\mathbb{Z}^d;\ell^2(\mathbb{N}))}, \tag{6.16}$$

with $\Theta_{\Sigma^d_{\leq l}}$ defined in (6.7). In particular, (6.16) implies the maximal estimate

$$\left\| \left(\sum_{\iota \in \mathbb{N}} \sup_{n \in \mathbb{N}} \big|T_{\mathbb{Z}^d}\big[\Theta_{\Sigma^d_{\leq l}}[\Upsilon^\Gamma_n \eta^{\Gamma^c}_{\leq m}]\big] f_\iota\big|^2 \right)^{1/2} \right\|_{\ell^p(\mathbb{Z}^d)} \leq C(l+1)\|f\|_{\ell^p(\mathbb{Z}^d;\ell^2(\mathbb{N}))}.$$

Some remarks about Theorem 6.14 are in order.

1. Theorem 6.14 is a semi-norm variant of the Ionescu–Wainger [34] theorem for oscillations. The proof below works also for r-variations or jumps in place of oscillations as well as for norms corresponding to real interpolation spaces. We refer to [50] for definitions.
2. In practice, Theorem 6.14 will be applied with $\Gamma = [d]$. However, the concept of Γ-lifted sequences is introduced here for further references.
3. A careful inspection of the proof below allows us to show that the conclusion of Theorem 6.14 also holds in $\mathbb{R}^d$. For every $d \in \mathbb{Z}_+$, every sequence $\Upsilon = (\Upsilon_n : \mathbb{R}^d \to \mathbb{C} : n \in \mathbb{Z})$ of measurable functions satisfying conditions (6.10), (6.11), (6.12) and (6.13), and for every $p \in (1, \infty)$, there exists a constant $C > 0$ such that for every $f = (f_\iota : \iota \in \mathbb{N}) \in L^p(\mathbb{R}^d; \ell^2(\mathbb{N}))$, one has

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{Z})} \left\| \left(\sum_{\iota \in \mathbb{N}} O_{I,J}\big(T_{\mathbb{R}^d}[\Upsilon_n] f_\iota : n \in \mathbb{Z}\big)^2 \right)^{1/2} \right\|_{L^p(\mathbb{R}^d)} \leq C\|f\|_{L^p(\mathbb{R}^d;\ell^2(\mathbb{N}))}. \tag{6.17}$$

An important feature of our approach is that we do not need to invoke the corresponding inequality for martingales in the proof. This stands in a sharp contrast to variants of inequality (6.17) involving r-variations, where all arguments to the best of our knowledge use the corresponding r-variational inequalities for martingales.

Proof of Theorem 6.14. Fix $p \in (1, \infty)$ and a sequence $f = (f_\iota : \iota \in \mathbb{N}) \in \ell^2(\mathbb{Z}^d; \ell^2(\mathbb{N})) \cap \ell^p(\mathbb{Z}^d; \ell^2(\mathbb{N}))$. For each $l \in \mathbb{N}$, define an integer

$$\kappa_l := \left\lfloor \big(100C_\rho + \log_2(\delta_\Upsilon \log_2 \tau)^{-1}\big)(l+1) \right\rfloor + 2, \tag{6.18}$$

where C_ρ is the constant from Remark 6.1; see property (iv). By (2.17), it only suffices to establish (6.16), which will follow from the oscillation inequalities, respectively, for small scales

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{N}_{<2^{\kappa_l}})} \left\| \left(\sum_{\iota \in \mathbb{N}} O_{I,J}\big(T_{\mathbb{Z}^d}\big[\Theta_{\Sigma^d_{\leq l}}[\Upsilon^\Gamma_n \eta^{\Gamma^c}_{\leq m}]\big] f_\iota : n \in \mathbb{N}_{<2^{\kappa_l}}\big)^2 \right)^{1/2} \right\|_{\ell^p(\mathbb{Z}^d)} \lesssim (l+1)\|f\|_{\ell^p(\mathbb{Z}^d;\ell^2(\mathbb{N}))}, \tag{6.19}$$

and large scales

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{N}_{\geq 2^{\kappa_l}})} \left\| \left(\sum_{\iota \in \mathbb{N}} O_{I,J}\big(T_{\mathbb{Z}^d}\big[\Theta_{\Sigma^d_{\leq l}}[\Upsilon^\Gamma_n \eta^{\Gamma^c}_{\leq m}]\big] f_\iota : n \in \mathbb{N}_{\geq 2^{\kappa_l}}\big)^2 \right)^{1/2} \right\|_{\ell^p(\mathbb{Z}^d)} \lesssim \|f\|_{\ell^p(\mathbb{Z}^d;\ell^2(\mathbb{N}))}. \tag{6.20}$$

Step 1. We now prove inequality (6.19). We fix $J \in \mathbb{Z}_+$ and a sequence $I \in \mathfrak{S}_J(\mathbb{N}_{<2^{\kappa_l}})$. Then, by the Rademacher–Menshov inequality (2.14), we see that

$$
\left\| \left(\sum_{\iota \in \mathbb{N}} O_{I,J} \big(T_{\mathbb{Z}^d} \big[\Theta_{\Sigma_{\leq l}^d} [\Upsilon_n^\Gamma \eta_{\leq m}^{\Gamma c}] \big] f_\iota : n \in \mathbb{N}_{<2^{\kappa_l}} \big)^2 \right)^{1/2} \right\|_{\ell^p(\mathbb{Z}^d)}
$$
$$
\lesssim \sum_{v=0}^{\kappa_l} \left\| \left(\sum_{\iota \in \mathbb{N}} \sum_{u=0}^{2^{\kappa_l-v}-1} \big| \sum_{n \in U_u^v} T_{\mathbb{Z}^d} \big[\Theta_{\Sigma_{\leq l}^d} [(\Upsilon_{n+1}^\Gamma - \Upsilon_n^\Gamma)\eta_{\leq m}^{\Gamma c}] \big] f_\iota \big|^2 \right)^{1/2} \right\|_{\ell^p(\mathbb{Z}^d)},
$$

where $U_u^v = [u2^v, (u+1)2^v) \cap \mathbb{Z}$. Hence, it suffices to prove

$$
\left\| \left(\sum_{\iota \in \mathbb{N}} \sum_{u=0}^{2^{\kappa_l-v}-1} \big| \sum_{n \in U_u^v} T_{\mathbb{Z}^d} \big[\Theta_{\Sigma_{\leq l}^d} [(\Upsilon_{n+1}^\Gamma - \Upsilon_n^\Gamma)\eta_{\leq m}^{\Gamma c}] \big] f_\iota \big|^2 \right)^{1/2} \right\|_{\ell^p(\mathbb{Z}^d)} \lesssim \|f\|_{\ell^p(\mathbb{Z}^d; \ell^2(\mathbb{N}))}, \tag{6.21}
$$

uniformly in v. By Theorem 6.5 and by our choice of κ_l in (6.18), since $m \leq -10C_\rho 2^{2\rho l}$, (6.21) will follow if for every sequence $(f_\iota : \iota \in \mathbb{N}) \in L^2(\mathbb{R}^d; \ell^2(\mathbb{N})) \cap L^p(\mathbb{R}^d; \ell^2(\mathbb{N}))$,

$$
\left\| \left(\sum_{\iota \in \mathbb{N}} \sum_{u=0}^{2^{\kappa_l-v}-1} \big| \sum_{n \in U_u^v} T_{\mathbb{R}^d} [(\Upsilon_{n+1}^\Gamma - \Upsilon_n^\Gamma)\eta_{\leq m}^{\Gamma c}] f_\iota \big|^2 \right)^{1/2} \right\|_{L^p(\mathbb{R}^d)} \lesssim \|f\|_{L^p(\mathbb{R}^d; \ell^2(\mathbb{N}))} \tag{6.22}
$$

holds uniformly in v.

To prove inequality (6.22), in view of Lemma 2.3, it suffices to show that for every $p \in (1, \infty)$ and for every $f \in L^p(\mathbb{R}^d)$, one has

$$
\sup_{(\omega_u) \in \{-1,1\}^{\mathbb{N}}} \left\| \sum_{u=0}^{2^{\kappa_l-v}-1} \omega_u \big[T_{\mathbb{R}^d} [(\Upsilon_{(u+1)2^v}^\Gamma - \Upsilon_{u2^v}^\Gamma)\eta_{\leq m}^{\Gamma c}] f \big] \right\|_{L^p(\mathbb{R}^d)} \lesssim_p \|f\|_{L^p(\mathbb{R}^d)}, \tag{6.23}
$$

uniformly in $v \in [0, \kappa_l]$ and l. The proof of (6.23), using conditions (6.10), (6.11), (6.12) and (6.13), follows from standard Littlewood–Paley theory as developed in [22]. We refer, for instance, to [51] for details in this context.

Step 2. We now prove inequality (6.20). By the support condition (6.15), we may write (see property (iv) from Remark 6.1)

$$
T_{\mathbb{Z}^d} \big[\Theta_{\Sigma_{\leq l}^d} [\Upsilon_n^\Gamma \eta_{\leq m}^{\Gamma c}] \big] = T_{\mathbb{Z}^d} \big[\Theta_{\Sigma_{\leq l}^d} [\Upsilon_n^\Gamma (1 - \eta_{\leq -2^{2C_\rho l}}^\Gamma)\eta_{\leq m}^{\Gamma c}] \big] + T_{\mathbb{Z}^d} \big[\Theta_{\Sigma_{\leq l}^d} [\Upsilon_n^\Gamma \eta_{\leq -2^{2C_\rho l}}^\Gamma \eta_{\leq m}^{\Gamma c}] \big],
$$

where $\eta_{\leq -2^{2C_\rho l}}^\Gamma := \prod_{i \in \Gamma} \eta_{\leq -2^{2C_\rho l}}^{(i)}$, (see definition (2.2)). The proof of (6.20) will be complete if we show (6.20) with $T_{\mathbb{Z}^d} \big[\Theta_{\Sigma_{\leq l}^d} [\Upsilon_n^\Gamma (1 - \eta_{\leq -2^{2C_\rho l}}^\Gamma)\eta_{\leq m}^{\Gamma c}] \big]$, and $T_{\mathbb{Z}^d} \big[\Theta_{\Sigma_{\leq l}^d} [\Upsilon_n^\Gamma \eta_{\leq -2^{2C_\rho l}}^\Gamma \eta_{\leq m}^{\Gamma c}] \big]$ in place of $T_{\mathbb{Z}^d} \big[\Theta_{\Sigma_{\leq l}^d} [\Upsilon_n^\Gamma \eta_{\leq m}^{\Gamma c}] \big]$. To establish (6.20) with $T_{\mathbb{Z}^d} \big[\Theta_{\Sigma_{\leq l}^d} [\Upsilon_n^\Gamma (1 - \eta_{\leq -2^{2C_\rho l}}^\Gamma)\eta_{\leq m}^{\Gamma c}] \big]$, it suffices to prove that for every $p \in (1, \infty)$, there exists $\delta_p \in (0, 1)$ such that for every $n \geq 2^{\kappa_l}$ and every $f = (f_\iota : \iota \in \mathbb{N}) \in \ell^2(\mathbb{Z}^d; \ell^2(\mathbb{N})) \cap \ell^p(\mathbb{Z}^d; \ell^2(\mathbb{N}))$, one has

$$
\left\| \left(\sum_{\iota \in \mathbb{N}} \big| T_{\mathbb{Z}^d} \big[\Theta_{\Sigma_{\leq l}^d} [\Upsilon_n^\Gamma (1 - \eta_{\leq -2^{2C_\rho l}}^\Gamma)\eta_{\leq m}^{\Gamma c}] \big] f_\iota \big|^2 \right)^{1/2} \right\|_{\ell^p(\mathbb{Z}^d)} \lesssim \tau^{-\delta_p n} \|f\|_{\ell^p(\mathbb{Z}^d; \ell^2(\mathbb{N}))}. \tag{6.24}
$$

Inequality (6.24), in view of Lemma 2.3 and Theorem 6.5, can be reduced to showing that for every $p \in (1, \infty)$, there exists $\delta_p \in (0, 1)$ such that

$$
\| T_{\mathbb{R}^d} [\Upsilon_n^\Gamma (1 - \eta_{\leq -2^{2C_\rho l}}^\Gamma)\eta_{\leq m}^{\Gamma c}] f \|_{L^p(\mathbb{R}^d)} \lesssim \tau^{-\delta_p n} \|f\|_{L^p(\mathbb{R}^d)} \tag{6.25}
$$

holds for every $n \geq 2^{\kappa_l}$. By interpolation, it suffices to prove (6.25) for $p = 2$ and by Plancherel's theorem, this reduces to showing that

$$|\Upsilon_n^\Gamma(\xi)(1 - \eta_{\leq -2^2 C_\rho l}^\Gamma(\xi))\eta_{\leq m}^{\Gamma^c}(\xi)| \lesssim \tau^{-\delta_\Upsilon n/2}$$

holds uniformly in ξ for all $n \geq 2^{\kappa_l}$. This follows from the definition of κ_l and (6.11).

Step 3. We now establish (6.20) with $T_{\mathbb{Z}^d}\big[\Theta_{\Sigma_{\leq l}^d}[\Upsilon_n^\Gamma \eta_{\leq -2^2 C_\rho l}^\Gamma \eta_{\leq m}^{\Gamma^c}]\big]$ in place of $T_{\mathbb{Z}^d}\big[\Theta_{\Sigma_{\leq l}^d}[\Upsilon_n^\Gamma \eta_{\leq m}^{\Gamma^c}]\big]$. Taking $Q_{\leq l}$ from property (iv), note that

$$T_{\mathbb{Z}^d}\big[\Theta_{\Sigma_{\leq l}^d}[\Upsilon_n^\Gamma \eta_{\leq -2^2 C_\rho l}^\Gamma \eta_{\leq m}^{\Gamma^c}]\big] = T_{\mathbb{Z}^d}\big[\Theta_{Q_{\leq l}^{-1}[Q_{\leq l}]^k}^\Gamma[\Upsilon_n^\Gamma \eta_{\leq -2^2 C_\rho l}^\Gamma]\big] T_{\mathbb{Z}^d}\big[\Theta_{\Sigma_{\leq l}^d}[\eta_{\leq m}^{[d]}]\big].$$

Using this factorization, it suffices to show that

$$\left\|\Big(\sum_{\iota \in \mathbb{N}}\big|T_{\mathbb{Z}^d}\big[\Theta_{\Sigma_{\leq l}^d}[\eta_{\leq m}^{[d]}]\big]f_\iota\big|^2\Big)^{1/2}\right\|_{\ell^p(\mathbb{Z}^d)} \lesssim \|f\|_{\ell^p(\mathbb{Z}^d;\ell^2(\mathbb{N}))}, \tag{6.26}$$

and

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{N}_{\geq 2^{\kappa_l}})}\left\|\Big(\sum_{\iota \in \mathbb{N}}O_{I,J}\big(T_{\mathbb{Z}^d}\big[\Theta_{Q_{\leq l}^{-1}[Q_{\leq l}]^k}^\Gamma[\Upsilon_n^\Gamma \eta_{\leq -2^2 C_\rho l}^\Gamma]\big]f_\iota : n \in \mathbb{N}_{\geq 2^{\kappa_l}}\big)^2\Big)^{1/2}\right\|_{\ell^p(\mathbb{Z}^d)}$$
$$\lesssim \|f\|_{\ell^p(\mathbb{Z}^d;\ell^2(\mathbb{N}))}. \tag{6.27}$$

By Lemma 2.3 and Theorem 6.5, the bound (6.26) follows from

$$\big\|T_{\mathbb{R}^d}[\eta_{\leq m}^{[d]}]f\big\|_{L^p(\mathbb{R}^d)} \lesssim_p \|f\|_{L^p(\mathbb{R}^d)},$$

which clearly holds for all $p \in [1, \infty]$. To prove (6.27), we can use the sampling principle formulated in Proposition 6.9 to reduce matters to proving

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{N}_{\geq 2^{\kappa_l}})}\left\|\Big(\sum_{\iota \in \mathbb{N}}O_{I,J}\big(T_{\mathbb{R}^k}[\Upsilon_n]f_\iota : n \in \mathbb{N}_{\geq 2^{\kappa_l}}\big)^2\Big)^{1/2}\right\|_{L^p(\mathbb{R}^k)} \lesssim \|f\|_{L^p(\mathbb{R}^k;\ell^2(\mathbb{N}))}. \tag{6.28}$$

To do this, we carefully choose the finite dimensional Banach spaces B_1 and B_2 in Proposition 6.9 to accommodate the oscillation semi-norm $O_{I,J}$. See the remark after Proposition 6.9.

Step 4. Let η be a smooth function with $\mathbb{1}_{[-1,1]^k} \leq \eta \leq \mathbb{1}_{[-\tau,\tau]^k}$ and set $\chi_n(\xi) := \eta(\tau^{-n} \circ \xi)$ Using conditions (6.10), (6.11), (6.12) and (6.13), we see that Theorem B in [22] implies

$$\left\|\Big(\sum_{n \in \mathbb{N}}\big|T_{\mathbb{R}^k}[\Upsilon_n - \chi_{-n}]f\big|^2\Big)^{1/2}\right\|_{L^p(\mathbb{R}^k)} \lesssim \|f\|_{L^p(\mathbb{R}^k)} \tag{6.29}$$

for $1 < p < \infty$ since $|\Upsilon_n(\xi) - \chi_{-n}(\xi)| \lesssim \min(|\tau^n \circ \xi|, |\tau^n \circ \xi|^{-1})^{\delta_\Upsilon}$ and both maximal functions $\sup_{n \in \mathbb{N}}|T_{\mathbb{R}^k}[\Upsilon_n]f|$ and $\sup_{n \in \mathbb{N}}|T_{\mathbb{R}^k}[\chi_{-n}]f|$ are both bounded on all $L^q(\mathbb{R}^k)$ for all $1 < q < \infty$.

Using Lemma 2.3, we see that inequality (6.29) reduces (6.28) to proving

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{N})}\left\|\Big(\sum_{\iota \in \mathbb{N}}O_{I,J}\big(T_{\mathbb{R}^k}[\chi_n]f_\iota : n \in \mathbb{Z}\big)^2\Big)^{1/2}\right\|_{L^p(\mathbb{R}^k)} \lesssim \|f\|_{L^p(\mathbb{R}^k;\ell^2(\mathbb{N}))}. \tag{6.30}$$

To prove (6.30), we note that for every $m < n$, we have

$$\chi_m \chi_n = \chi_m.$$

We fix $J \in \mathbb{Z}_+$ and a sequence $I \in \mathfrak{S}_J(\mathbb{N})$. Then

$$
O_{I,J}\left(T_{\mathbb{R}^k}[\chi_n]f_\iota : n \in \mathbb{N}\right) \lesssim \left(\sum_{j=0}^{J-1} \sup_{I_j \leq n < I_{j+1}} \left|T_{\mathbb{R}^k}[\chi_n - \chi_{I_j}]f_\iota\right|^2\right)^{1/2}
$$

$$
= \left(\sum_{j=0}^{J-1} \sup_{I_j < n < I_{j+1}} \left|T_{\mathbb{R}^k}[\chi_n]T_{\mathbb{R}^k}[\chi_{I_{j+1}} - \chi_{I_j}]f_\iota\right|^2\right)^{1/2}
$$

$$
\leq \left(\sum_{j \in \mathbb{N}} \sup_{n \in \mathbb{Z}} \left(\varphi_n * \left|T_{\mathbb{R}^k}[\chi_{I_{j+1}} - \chi_{I_j}]f_\iota\right|\right)^2\right)^{1/2},
$$

where $\varphi_n(x) := |T_{\mathbb{R}^k}[\chi_n](x)|$. Using this estimate and the Fefferman-Stein vector-valued maximal function estimate (see [57]), we conclude that

$$
\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{N})} \left\|\left(\sum_{\iota \in \mathbb{N}} O_{I,J}\left(T_{\mathbb{R}^k}[\chi_n]f_\iota : n \in \mathbb{Z}\right)^2\right)^{1/2}\right\|_{L^p(\mathbb{R}^k)}
$$

$$
\leq \sup_{I \in \mathfrak{S}_\infty(\mathbb{N})} \left\|\left(\sum_{\iota \in \mathbb{N}} \sum_{j \in \mathbb{N}} \sup_{n \in \mathbb{Z}} \left(\varphi_n * \left|T_{\mathbb{R}^k}[\chi_{I_{j+1}} - \chi_{I_j}]f_\iota\right|\right)^2\right)^{1/2}\right\|_{L^p(\mathbb{R}^k)}
$$

$$
\lesssim_p \sup_{I \in \mathfrak{S}_\infty(\mathbb{N})} \left\|\left(\sum_{\iota \in \mathbb{N}} \sum_{j \in \mathbb{N}} \left|T_{\mathbb{R}^k}[\chi_{I_{j+1}} - \chi_{I_j}]f_\iota\right|^2\right)^{1/2}\right\|_{L^p(\mathbb{R}^k)}. \tag{6.31}
$$

As above, using Theorem B in [22], we see that for every $p \in (1, \infty)$,

$$
\sup_{I \in \mathfrak{S}_\infty(\mathbb{N})} \left\|\left(\sum_{j \in \mathbb{N}} \left|T_{\mathbb{R}^k}[\chi_{I_{j+1}} - \chi_{I_j}]f\right|^2\right)^{1/2}\right\|_{L^p(\mathbb{R}^k)} \lesssim_p \|f\|_{L^p(\mathbb{R}^k)}. \tag{6.32}
$$

Then invoking (6.32) and Lemma 2.3, we obtain

$$
\sup_{I \in \mathfrak{S}_\infty(\mathbb{N})} \left\|\left(\sum_{\iota \in \mathbb{N}} \sum_{j \in \mathbb{N}} \left|T_{\mathbb{R}^k}[\chi_{I_{j+1}} - \chi_{I_j}]f_\iota\right|^2\right)^{1/2}\right\|_{L^p(\mathbb{R}^k)} \lesssim_p \|f\|_{L^p(\mathbb{R}^k; \ell^2(\mathbb{N}))}. \tag{6.33}
$$

Combining (6.31) with (6.33), we obtain the desired claim in (6.30), and this completes the proof of Theorem 6.14. □

6.3. *Multi-parameter semi-norm variant of Theorem 6.5*

We will generalize Theorem 6.14 to the multi-parameter setting for a class of multipliers arising in our question. We formulate our main result in the two-parameter setting, but all arguments are adaptable to multi-parameter settings.

Let $P \in \mathbb{R}[m_1, m_2]$ be a polynomial with $\deg P \geq 2$ such that

$$
P(m_1, m_2) := \sum_{(\gamma_1, \gamma_2) \in S_P} c_{\gamma_1, \gamma_2} m_1^{\gamma_1} m_2^{\gamma_2}, \tag{6.34}
$$

where $c_{(0,0)} = 0$. In addition, we assume that P is non-degenerate in the sense that $S_P \cap (\mathbb{Z}_+ \times \mathbb{Z}_+) \neq \emptyset$; see the remark below Theorem 1.11. Let $r \in \mathbb{Z}_+$ be the number of vertices in the backwards Newton diagram N_P corresponding to the polynomial P from (6.34). For any vertex $v_j = (v_{j,1}, v_{j,2})$ of N_P, we denote the associated monomial by

$$
P^j(m_1, m_2) := c_{(v_{j,1}, v_{j,2})} m_1^{v_{j,1}} m_2^{v_{j,2}}. \tag{6.35}
$$

From Section 4 (see Remark 4.5), we know that P^j is the main monomial in the sector $S(j)$ for $j \in [r]$.

We fix the lacunarity factor $\tau > 1$. Throughout this subsection, we allow all the implied constants to depend on τ. For real numbers $M_1, M_2 \geq 1$ and $\xi \in \mathbb{R}$, we consider the multiplier

$$\mathfrak{m}^P_{M_1,M_2}(\xi) := \frac{1}{(1-\tau^{-1})^2} \int_{\tau^{-1}}^1 \int_{\tau^{-1}}^1 e(P_\xi(M_1 y_1, M_2 y_2)) dy_1 dy_2, \qquad (6.36)$$

where recall $P_\xi \in \mathbb{R}[\mathrm{m}_1, \mathrm{m}_2]$ is defined as $P_\xi(\mathrm{m}_1, \mathrm{m}_2) = \xi P(\mathrm{m}_1, \mathrm{m}_2)$.

As an application of Theorem 6.14, we obtain the following two-parameter oscillation inequality.

Theorem 6.37. *Let $\tau > 1$ be given and let $(\mathfrak{m}^P_{M_1,M_2} : (M_1, M_2) \in \mathbb{D}_\tau \times \mathbb{D}_\tau)$ be the two-parameter sequence of multipliers from (6.36) corresponding to the polynomial P from (6.34). Let $r \in \mathbb{Z}_+$ be the number of vertices in the backwards Newton diagram N_P. For every $\rho \in (0, 1)$ and $p \in (1, \infty)$ and any $j \in [r]$, there exists an absolute constant $0 < C = C(p, \rho, \tau, j, P) < \infty$ such that for every integers $l \in \mathbb{N}$ and $m \leq -10 C_\rho 2^{2\rho l}$ and for every $f = (f_\iota : \iota \in \mathbb{N}) \in \ell^p(\mathbb{Z}; \ell^2(\mathbb{N}))$, one has*

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{S}_\tau(j))} \left\| \left(\sum_{\iota \in \mathbb{N}} O_{I,J}\left(T_{\mathbb{Z}}\left[\Theta_{\Sigma_{\leq l}}[\mathfrak{m}^P_{M_1,M_2} \eta_{\leq m}]\right] f_\iota : (M_1, M_2) \in \mathbb{S}_\tau(j)\right)^2 \right)^{1/2} \right\|_{\ell^p(\mathbb{Z})} \qquad (6.38)$$
$$\leq C(l+1)\|f\|_{\ell^p(\mathbb{Z}; \ell^2(\mathbb{N}))},$$

with $\Theta_{\Sigma_{\leq l}}$ defined in (6.7). In particular, (6.38) also implies the maximal estimate

$$\left\| \left(\sum_{\iota \in \mathbb{N}} \sup_{(M_1, M_2) \in \mathbb{S}_\tau(j)} \left| T_{\mathbb{Z}}\left[\Theta_{\Sigma_{\leq l}}[\mathfrak{m}^P_{M_1,M_2} \eta_{\leq m}]\right] f_\iota \right|^2 \right)^{1/2} \right\|_{\ell^p(\mathbb{Z})} \leq C(l+1)\|f\|_{\ell^p(\mathbb{Z}; \ell^2(\mathbb{N}))}.$$

Some remarks about Theorem 6.37 are in order.

1. Theorem 6.37 is the simplest instance of a multi-parameter oscillation variant of the Ionescu–Wainger theorem [34]. More general variants of Theorem 6.37 can be also proved. For instance, an analogue of Theorem 6.37 for the following multipliers

$$\mathfrak{m}^P_{M_1,M_2}(\xi_1, \xi_2, \xi_3) = \int_0^1 \int_0^1 e(\xi_1(M_1 y_1) + \xi_2(M_2 y_2) + \xi_3 P(M_1 y_1, M_2 y_2)) dy_1 dy_2$$

   can be established using the methods of the paper. However, this goes beyond the scope of this paper and will be discussed in the future.

2. In contrast to the one-parameter theory, it is not clear whether multi-parameter r-variational or jump counterparts of Theorem 6.37 are available. As far as we know, it is not even clear if there are useful multi-parameter definitions of r-variational or jump semi-norms. From this point of view, the multi-parameter oscillation semi-norm is an invaluable tool allowing us to handle pointwise convergence problems in the multi-parameter setting.

3. A careful inspection of the proof allows us to establish an analogue of Theorem 6.37 in the continuous setting. Namely, for every $p \in (1, \infty)$, there is a constant $C > 0$ such that for every $f = (f_\iota : \iota \in \mathbb{N}) \in L^p(\mathbb{R}; \ell^2(\mathbb{N}))$, one has

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{S}_\tau(j))} \left\| \left(\sum_{\iota \in \mathbb{N}} O_{I,J}\left(T_{\mathbb{R}}[\mathfrak{m}^P_{M_1,M_2}] f_\iota : (M_1, M_2) \in \mathbb{S}_\tau(j)\right)^2 \right)^{1/2} \right\|_{L^p(\mathbb{R})} \leq C\|f\|_{L^p(\mathbb{R}; \ell^2(\mathbb{N}))}.$$

Proof of Theorem 6.37. We will only prove Theorem 6.37 for $j = r = 1$ or for $1 \leq j < r$ with $r \geq 2$. The same argument can be used to prove the case for $j = r$. In view of (2.17), it suffices to prove (6.38). We divide the proof into two steps to make the argument clearer.

Step 1. We prove that for every $p \in (1, \infty)$ and every $f = (f_\iota : \iota \in \mathbb{N}) \in \ell^P(\mathbb{Z}; \ell^2(\mathbb{N}))$, one has

$$\left\| \left(\sum_{(M_1, M_2) \in \mathbb{S}_\tau(j)} \sum_{\iota \in \mathbb{N}} \left| T_\mathbb{Z}\big[\Theta_{\Sigma_{\leq l}}\big[(\mathfrak{m}^P_{M_1, M_2} - \mathfrak{m}^{P^j}_{M_1, M_2})\eta_{\leq m}\big]\big] f_\iota \right|^2 \right)^{1/2} \right\|_{\ell^P(\mathbb{Z})} \lesssim \|f\|_{\ell^P(\mathbb{Z}; \ell^2(\mathbb{N}))}.$$

Using (4.14) and (4.15), it suffices to prove that for every $p \in (1, \infty)$, there is $\sigma_{j,p} \in (0, 1)$ such that for every $N \in \mathbb{N}$, $i \in [2]$ and every $f = (f_\iota : \iota \in \mathbb{N}) \in \ell^P(\mathbb{Z}; \ell^2(\mathbb{N}))$, one has

$$\left\| \left(\sum_{(M_1, M_2) \in \mathbb{S}^N_{\tau, i}(j)} \sum_{\iota \in \mathbb{N}} \left| T_\mathbb{Z}\big[\Theta_{\Sigma_{\leq l}}\big[(\mathfrak{m}^P_{M_1, M_2} - \mathfrak{m}^{P^j}_{M_1, M_2})\eta_{\leq m}\big]\big] f_\iota \right|^2 \right)^{1/2} \right\|_{\ell^P(\mathbb{Z})} \lesssim \tau^{-\sigma_{j,p} N} \|f\|_{\ell^P(\mathbb{Z}; \ell^2(\mathbb{N}))}. \tag{6.39}$$

We only prove (6.39) for $i = 1$, as the proof for $i = 2$ is the same. By the construction of the sets $\mathbb{S}^N_{\tau, 1}(j)$ (see definition (4.15)), the problem becomes a one-parameter problem. Indeed, if $(M_1, M_2) \in \mathbb{S}^N_{\tau, 1}(j)$, then $(M_1, M_2) = (\tau^{n_1}, \tau^{n_2})$ and

$$(n_1, n_2) = \frac{n}{d_j}\omega_{j-1} + \frac{N}{d_j}(\omega_j + \omega_{j-1}) \quad \text{for some} \quad n \in \mathbb{Z}_+.$$

Defining $(n_1^k, n_2^k) := \frac{k}{d_j}\omega_{j-1} + \frac{N}{d_j}(\omega_j + \omega_{j-1})$ for any $k \in \mathbb{Z}_+$, inequality (6.39) can be written as

$$\left\| \left(\sum_{k \in \mathbb{Z}_+} \sum_{\iota \in \mathbb{N}} \left| T_\mathbb{Z}\big[\Theta_{\Sigma_{\leq l}}\big[(\mathfrak{m}^P_{\tau^{n_1^k}, \tau^{n_2^k}} - \mathfrak{m}^{P^j}_{\tau^{n_1^k}, \tau^{n_2^k}})\eta_{\leq m}\big]\big] f_\iota \right|^2 \right)^{1/2} \right\|_{\ell^P(\mathbb{Z})} \lesssim \tau^{-\sigma_{j,p} N} \|f\|_{\ell^P(\mathbb{Z}; \ell^2(\mathbb{N}))}.$$

By Lemma 2.3 and Theorem 6.5, it suffices to prove that for every $p \in (1, \infty)$, there is $\sigma_{j,p} \in (0, 1)$ such that for every $N \in \mathbb{N}$ and $f \in L^P(\mathbb{R})$, one has

$$\sup_{(\varepsilon_k : k \in \mathbb{Z}_+) \in \{0, 1\}^{\mathbb{Z}_+}} \left\| \sum_{k \in \mathbb{Z}_+} \varepsilon_k T_\mathbb{R}\big[(\mathfrak{m}^P_{\tau^{n_1^k}, \tau^{n_2^k}} - \mathfrak{m}^{P^j}_{\tau^{n_1^k}, \tau^{n_2^k}})\big] f \right\|_{L^P(\mathbb{R})} \lesssim \tau^{-\sigma_{j,p} N} \|f\|_{L^P(\mathbb{R})}. \tag{6.40}$$

By (6.34), (6.35) and Lemma 4.10, we obtain

$$
\begin{aligned}
|P(\tau^{n_1} y_1, \tau^{n_2} y_2) - P^j(\tau^{n_1} y_1, \tau^{n_2} y_2)| &\leq \sum_{(\gamma_1, \gamma_2) \in S_P \setminus \{v_j\}} |c_{\gamma_1, \gamma_2}| \tau^{(\gamma_1, \gamma_2) \cdot (n_1, n_2)} |y_1|^{\gamma_1} |y_2|^{\gamma_2} \\
&\leq \big(\sup_{v \in S_P} |c_v| \big) \tau^{(n_1, n_2) \cdot v_j} \sum_{v \in S_P \setminus \{v_j\}} \tau^{(n_1, n_2) \cdot (v - v_j)} \\
&\leq \#S_P \big(\sup_{v \in S_P} |c_v| \big) \tau^{(n_1, n_2) \cdot v_j} \tau^{-\sigma_j N}
\end{aligned}
$$

whenever $|y_1|, |y_2| \leq 1$, with $\sigma_j > 0$ defined in (4.11). Consequently, we have

$$|\mathfrak{m}^P_{\tau^{n_1^k}, \tau^{n_2^k}}(\xi) - \mathfrak{m}^{P^j}_{\tau^{n_1^k}, \tau^{n_2^k}}(\xi)| \lesssim_P \tau^{-\sigma_j N} (\tau^{(n_1^k, n_2^k) \cdot v_j} |\xi|). \tag{6.41}$$

Moreover, by van der Corput's lemma (Proposition 2.6), we can find a $\delta_0 \in (0, 1)$ such that

$$|\mathfrak{m}^P_{\tau^{n_1^k}, \tau^{n_2^k}}(\xi) - \mathfrak{m}^{P^j}_{\tau^{n_1^k}, \tau^{n_2^k}}(\xi)| \lesssim_P (\tau^{(n_1^k, n_2^k) \cdot v_j} |\xi|)^{-\delta_0} \tag{6.42}$$

for sufficiently large $N \in \mathbb{N}$. A convex combination of (6.41) and (6.42) gives

$$|\mathfrak{m}^{P}_{\tau^{n^k_1}, \tau^{n^k_2}}(\xi) - \mathfrak{m}^{P^j}_{\tau^{n^k_1}, \tau^{n^k_2}}(\xi)| \lesssim_P \tau^{-\sigma'_j N} \min\left\{(\tau^{(n^k_1, n^k_2) \cdot v_j}|\xi|)^{\delta'_0}, (\tau^{(n^k_1, n^k_2) \cdot v_j}|\xi|)^{-\delta'_0}\right\}, \qquad (6.43)$$

for some $\delta'_0, \sigma'_j \in (0, 1)$.

Using (6.43) and Plancherel's theorem, we obtain (6.40) for $p = 2$. Standard Littlewood–Paley theory arguments (see, for example, Theorem D in [22]) allows us then to obtain (6.40) for all $p \in (1, \infty)$.

Step 2. The argument from the first step allows us to reduce matters to proving

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}^2_\tau)} \left\| \left(\sum_{\iota \in \mathbb{N}} O_{I,J}\big(T_{\mathbb{Z}}\big[\Theta_{\Sigma \leq l}[\mathfrak{m}^{P^j}_{M_1, M_2} \eta_{\leq m}]\big] f_\iota : (M_1, M_2) \in \mathbb{D}^2_\tau \big)^2 \right)^{1/2} \right\|_{\ell^P(\mathbb{Z})}$$

$$\lesssim (l + 1)\|f\|_{\ell^P(\mathbb{Z}; \ell^2(\mathbb{N}))}.$$

We define a new one-parameter multiplier

$$\mathfrak{g}^{P^j}_M(\xi) := \frac{1}{(1 - \tau^{-1})^2} \int_{\tau^{-1}}^{1} \int_{\tau^{-1}}^{1} e(c_{(v_{j,1}, v_{j,2})} M\xi y_1^{v_{j,1}} y_2^{v_{j,2}}) dy_1 dy_2.$$

Observe that by Theorem 6.14, we obtain

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}^2_\tau)} \left\| \left(\sum_{\iota \in \mathbb{N}} O_{I,J}\big(T_{\mathbb{Z}}\big[\Theta_{\Sigma \leq l}[\mathfrak{m}^{P^j}_{M_1, M_2} \eta_{\leq m}]\big] f_\iota : (M_1, M_2) \in \mathbb{D}^2_\tau \big)^2 \right)^{1/2} \right\|_{\ell^P(\mathbb{Z})}$$

$$\leq \sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{D}_\tau)} \left\| \left(\sum_{\iota \in \mathbb{N}} O_{I,J}\big(T_{\mathbb{Z}}\big[\Theta_{\Sigma \leq l}\mathfrak{g}^{P^j}_M \eta_{\leq m}\big] f_\iota : M \in \mathbb{D}_\tau \big)^2 \right)^{1/2} \right\|_{\ell^P(\mathbb{Z})} \lesssim (l + 1)\|f\|_{\ell^P(\mathbb{Z}; \ell^2(\mathbb{N}))}.$$

This completes the proof of the theorem. $\qquad \square$

7. Two-parameter circle method: Proof of Theorem 4.16

Throughout this section, $\tau > 1$ is fixed, and we allow all the implied constants to depend on τ. Let $P \in \mathbb{Z}[\mathrm{m}_1, \mathrm{m}_2]$ be a polynomial obeying $P(0, 0) = 0$, which is non-degenerate in the sense that $S_P \cap (\mathbb{Z}_+ \times \mathbb{Z}_+) \neq \emptyset$; see (1.14). For every real number $N \geq 1$, define

$$\chi_N(x) := \frac{1}{|(\tau^{-1}N, N] \cap \mathbb{Z}|} \mathbb{1}_{(\tau^{-1}N, N]}(x), \quad x \in \mathbb{R}.$$

For every real number $M_1, M_2 \geq 1$ and $\xi \in \mathbb{R}$, we consider the multiplier

$$m_{M_1, M_2}(\xi) := \sum_{m_1 \in \mathbb{Z}} \sum_{m_2 \in \mathbb{Z}} e(P_\xi(m_1, m_2)) \chi_{M_1}(m_1) \chi_{M_2}(m_2),$$

with $P_\xi(m_1, m_2) = \xi P(m_1, m_2)$. The corresponding partial multipliers are defined by

$$m^1_{m_1, M_2}(\xi) := \sum_{m_2 \in \mathbb{Z}} e(P_\xi(m_1, m_2)) \chi_{M_2}(m_2), \qquad m_1 \in \mathbb{Z},$$

$$m^2_{M_1, m_2}(\xi) := \sum_{m_1 \in \mathbb{Z}} e(P_\xi(m_1, m_2)) \chi_{M_1}(m_1), \qquad m_2 \in \mathbb{Z}. \tag{7.1}$$

We fix further notation and terminology. For functions $G : \mathbb{Q} \cap \mathbb{T} \to \mathbb{C}$, $\mathfrak{m} : \mathbb{T} \to \mathbb{C}$, a finite set $\Sigma \subset \mathbb{Q} \cap \mathbb{T}$, any $n \in \mathbb{Z}$ and any $\xi \in \mathbb{T}$, we define the following 1-periodic multiplier:

$$\Phi^{\Sigma}_{\leq n}[G, \mathfrak{m}](\xi) := \sum_{a/q \in \Sigma} G(a/q) \mathfrak{m}(\xi - a/q) \eta_{\leq n}(\xi - a/q). \tag{7.2}$$

In a similar way, for any $l \in \mathbb{N}$, $n \in \mathbb{Z}$, any $\xi \in \mathbb{T}$, we define the following projection multipliers (recall the definition of $\Sigma_{\leq l} := \Sigma^1_{\leq l}$ from (6.3))

$$\Delta_{\leq l, \leq n}(\xi) := \sum_{a/q \in \Sigma_{\leq l}} \eta_{\leq n}(\xi - a/q), \qquad \text{and} \qquad \Delta^c_{\leq l, \leq n}(\xi) := 1 - \Delta_{\leq l, \leq n}(\xi).$$

All these multipliers will be applied with different choices of parameters. For $\beta > 0$, $M_1, M_2, M > 0$, $N \geq 0$, and $v = (v_1, v_2) \in \mathbb{Z}^2$, we define

$$l^{\beta}(M) := \log_2\left((\log_{\tau} M)^{\beta}\right), \qquad \text{and} \qquad n^v_{M_1, M_2}(N) := \log_2(M_1^{v_1} M_2^{v_2}) - N. \tag{7.3}$$

Using (7.3), we also set

$$n^{v, \beta}_{M_1, M_2}(M) := n^v_{M_1, M_2}(l^{\beta}(M)) = \log_2(M_1^{v_1} M_2^{v_2} (\log_{\tau} M)^{-\beta}). \tag{7.4}$$

Definitions (7.3) and (7.4) will be applied with $v \in \mathbb{Z}^2$ being a vertex of the backwards Newton diagram N_P. In this section, we shall abbreviate $\mathfrak{m}^P_{M_1, M_2}$ to

$$\mathfrak{m}_{M_1, M_2}(\xi) := \frac{1}{(1 - \tau^{-1})^2} \int_{\tau^{-1}}^1 \int_{\tau^{-1}}^1 e(P_{\xi}(M_1 y_1, M_2 y_2)) dy_1 dy_2, \qquad \xi \in \mathbb{R}.$$

We also define the following two partial multipliers:

$$\begin{aligned}
\mathfrak{m}^1_{m_1, M_2}(\xi) &:= \frac{1}{1 - \tau^{-1}} \int_{\tau^{-1}}^1 e(P_{\xi}(m_1, M_2 y_2)) dy_2, \qquad \xi \in \mathbb{R}, \ m_1 \in \mathbb{Z}, \\
\mathfrak{m}^2_{M_1, m_2}(\xi) &:= \frac{1}{1 - \tau^{-1}} \int_{\tau^{-1}}^1 e(P_{\xi}(M_1 y_1, m_2)) dy_1, \qquad \xi \in \mathbb{R}, \ m_2 \in \mathbb{Z}.
\end{aligned} \tag{7.5}$$

Our main result of this section is Theorem 7.6, which is a restatement of Theorem 4.16.

Theorem 7.6. *Let $r \in \mathbb{Z}_+$ be the number of vertices in the backwards Newton diagram N_P. Then for every $p \in (1, \infty)$ and $j \in [r]$ and for every $f \in \ell^p(\mathbb{Z})$, one has*

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{S}_{\tau}(j))} \|O_{I, J}(T_{\mathbb{Z}}[m_{M_1, M_2}] f : (M_1, M_2) \in \mathbb{S}_{\tau}(j))\|_{\ell^p(\mathbb{Z})} \lesssim_{p, \tau} \|f\|_{\ell^p(\mathbb{Z})}. \tag{7.7}$$

The proof of Theorem 7.6 is divided into several steps. We apply iteratively the classical circle method, taking into account the geometry of the backwards Newton diagram N_P.

7.1. Preliminaries

The number of vertices $r \in \mathbb{Z}_+$ in the backwards Newton diagram N_P is fixed. Let $v_j = (v_{j,1}, v_{j,2})$ denote the vertex of N_P corresponding to $j \in [r]$.

It suffices to establish inequality (7.7) for $j = r = 1$ assuming additionally that $\log M_1 \leq \log M_2$ when $(M_1, M_2) \in \mathbb{S}_{\tau}(1)$, or for any $r \geq 2$ and any $1 \leq j < r$. Both cases ensure that

$$\log M_1 \lesssim \log M_2 \quad \text{whenever} \quad (M_1, M_2) \in \mathbb{S}_{\tau}(j), \tag{7.8}$$

which means that $M_1 \leq M_2^{K_j}$ for some $K_j > 0$; see Remark 4.5. The case when $j = r$ with $r \geq 2$ can be proved in much the same way, with the difference that $\log M_1 \gtrsim \log M_2$ whenever $(M_1, M_2) \in \mathbb{S}_\tau(r)$. We only outline the most important changes, omitting the details, which can be easily adjusted using the arguments below.

From now on, $p \in (1, \infty)$ is fixed and we let $p_0 \in (1, 2)$ be such that $p \in (p_0, p_0')$. The proof will involve several parameters that have to be suitably adjusted to $p \in (p_0, p_0')$.

We begin by setting

$$\theta_p := \left(\frac{1}{p_0} - \frac{1}{\min\{p, p'\}} \right) \left(\frac{1}{p_0} - \frac{1}{2} \right)^{-1} \in (0, 1).$$

We will take

$$\alpha > 100 \, \theta_p^{-1}, \qquad \text{and} \qquad \beta > 1000 \max \left\{ \delta^{-1}, (1 + \deg P)^5 \right\} (\alpha + 1), \tag{7.9}$$

where $\beta \in \mathbb{Z}_+$ plays the role of the parameter $\beta \in \mathbb{Z}_+$ from Proposition 5.37, and $\delta \in (0, 1)$ is the parameter that arises in the complete sum estimates; see Proposition 5.46.

Finally, we need the parameter $\rho > 0$, introduced in the Ionescu–Wainger multiplier theorem (see Theorem 6.5 as well as Theorem 6.14 and Theorem 6.37), to satisfy

$$\rho \beta < \frac{1}{1000}. \tag{7.10}$$

7.2. Minor arc estimates

We first establish the minor arcs estimates.

Claim 7.11. *For every $1 \leq j < r$ and for every $(M_1, M_2) \in \mathbb{S}_\tau(j)$, one has*

$$\|T_{\mathbb{Z}}[m_{M_1, M_2} \Delta^c_{\leq l^\beta(M_2), \leq -n_{M_1, M_2}^{v_j, \beta}(M_2)}] f\|_{\ell^2(\mathbb{Z})} \lesssim_\tau (\log M_2)^{-\alpha} \|f\|_{\ell^2(\mathbb{Z})}, \qquad f \in \ell^2(\mathbb{Z}), \tag{7.12}$$

with α as in (7.9). The same estimate holds when $j = r = 1$, as long as $\log M_1 \leq \log M_2$.

The case $j = r \geq 2$ requires a minor modification. Keeping in mind that $\log M_2 \lesssim \log M_1$, it suffices to establish an analogue of (7.12). Namely, one has

$$\|T_{\mathbb{Z}}[m_{M_1, M_2} \Delta^c_{\leq l^\beta(M_1), \leq -n_{M_1, M_2}^{v_j, \beta}(M_1)}] f\|_{\ell^2(\mathbb{Z})} \lesssim_\tau (\log M_1)^{-\alpha} \|f\|_{\ell^2(\mathbb{Z})}, \qquad f \in \ell^2(\mathbb{Z}).$$

Proof of Claim 7.11. Since $\log M_1 \lesssim \log M_2$, one has $\log M_{r,j}^* \simeq \log M_2$, where $M_{r,j}^*$ was defined in (5.36). We can also assume that M_2 is a large number. To prove (7.12), by Plancherel's theorem, it suffices to show for every $\xi \in \mathbb{T}$ that

$$\left| m_{M_1, M_2}(\xi) \Delta^c_{\leq l^\beta(M_2), \leq -n_{M_1, M_2}^{v_j, \beta}(M_2)}(\xi) \right| \lesssim (\log M_2)^{-\alpha}. \tag{7.13}$$

For this purpose, we use Dirichlet's principle to find a rational fraction a_0/q_0 such that $(a_0, q_0) = 1$ and $1 \leq q_0 \leq C M_1^{v_{j,1}} M_2^{v_{j,2}} \log(M_{r,j}^*)^{-\beta} = C 2^{n_{M_1, M_2}^{v_j, \beta}(M_{r,j}^*)}$ and

$$\left| \xi - \frac{a_0}{q_0} \right| \leq \frac{\log(M_{r,j}^*)^\beta}{q_0 C M_1^{v_{j,1}} M_2^{v_{j,2}}} \leq \frac{1}{q_0^2}$$

for a large constant $C > 1$ to be specified later. If $q_0 < \log(M_2)^\beta$, then $a_0/q_0 \in \Sigma_{\leq l^\beta(M_2)}$ and consequently the left-hand side of (7.13) vanishes (if $C > 1$ is large enough) and there is nothing to prove. Thus, we can assume that $\log(M^*_{r,j})^\beta \lesssim q_0 \lesssim M_1^{v_{j,1}} M_2^{v_{j,2}} \log(M^*_{r,j})^{-\beta}$. We now can apply Proposition 5.37 and obtain (7.13) as claimed. □

7.3. Major arcs estimates

Recalling (7.2), we begin with a simple approximation formula.

Lemma 7.14. *Suppose that $1 \leq j < r$ and $(M_1, M_2) \in \mathbb{S}_\tau(j)$. Then for every $0 \leq l, l' \leq l^\beta(M_2)$ and $(M'_1, M'_2) \in \mathbb{S}_\tau(j)$ and $m_1 \simeq M'_1$ such that $1 \leq M'_1 \leq M_1$ and $2^{C_\rho 2^{\rho l}} \leq M'_2 \leq M_2$, one has*

$$m^1_{m_1,M'_2}(\xi)\Delta_{\leq l, \leq -n^{v_j}_{M_1,M_2}(l')}(\xi) = \Phi^{\Sigma_{\leq l}}_{\leq -n^{v_j}_{M_1,M_2}(l')}[G^1_{m_1}, \mathfrak{m}^1_{m_1,M'_2}](\xi) + O(2^{C_\rho 2^{\rho l}}(M'_2)^{-1}), \qquad (7.15)$$

where $n^v_{M_1,M_2}(N)$, $G^1_{m_1}$, $m^1_{m_1,M_2}$ and $\mathfrak{m}^1_{m_1,M_2}$ were defined respectively in (7.4), (5.45), (7.1) and (7.5). In particular, (7.15) immediately yields

$$\begin{aligned}
m_{M'_1,M'_2}(\xi)\Delta_{\leq l, \leq -n^{v_j}_{M_1,M_2}(l')}(\xi) &= \sum_{m_1 \in \mathbb{Z}} \Phi^{\Sigma_{\leq l}}_{\leq -n^{v_j}_{M_1,M_2}(l')}[G^1_{m_1}, \mathfrak{m}^1_{m_1,M'_2}](\xi)\chi_{M'_1}(m_1) \\
&\quad + O(2^{C_\rho 2^{\rho l}}(M'_2)^{-1}).
\end{aligned} \qquad (7.16)$$

The same claims hold when $j = r = 1$, as long as $\log M_1 \leq \log M_2$.

A similar conclusion holds when $j = r \geq 2$. Taking into account that $\log M_2 \lesssim \log M_1$ whenever $(M_1, M_2) \in \mathbb{S}_\tau(r)$ and assuming that $0 \leq l, l' \leq l^\beta(M_1)$, one has for every $(M'_1, M'_2) \in \mathbb{S}_\tau(j)$ and $m_2 \simeq M'_2$ satisfying $2^{C_\rho 2^{\rho l}} \leq M'_1 \leq M_1$ and $1 \leq M'_2 \leq M_2$ that

$$m^2_{M'_1,m_2}(\xi)\Delta_{\leq l, \leq -n^{v_j}_{M_1,M_2}(l')}(\xi) = \Phi^{\Sigma_{\leq l}}_{\leq -n^{v_j}_{M_1,M_2}(l')}[G^2_{m_2}, \mathfrak{m}^2_{M'_1,m_2}](\xi) + O(2^{C_\rho 2^{\rho l}}(M'_1)^{-1}). \qquad (7.17)$$

In particular, (7.17) yields

$$m_{M'_1,M'_2}(\xi)\Delta_{\leq l, \leq -n^{v_j}_{M_1,M_2}(l')}(\xi) = \sum_{m_2 \in \mathbb{Z}} \Phi^{\Sigma_{\leq l}}_{\leq -n^{v_j}_{M_1,M_2}(l')}[G^2_{m_2}, \mathfrak{m}^2_{M'_1,m_2}](\xi)\chi_{M'_2}(m_2) + O(2^{C_\rho 2^{\rho l}}(M'_1)^{-1}).$$

Proof of Lemma 7.14. For every $a/q \in \Sigma_{\leq l}$, we note

$$e(P_\xi(m_1, m_2)) = e(P_{\xi-a/q}(m_1, qm + r_2))e(P_{a/q}(m_1, r_2)), \qquad (7.18)$$

whenever $m_1 \in \mathbb{Z}$, $m_2 = qm + r_2$ and $r_2 \in \mathbb{Z}_q$. Then, by (7.18), since $q \leq 2^{C_\rho 2^{\rho l}} \leq M'_2$, we have

$$\sum_{m_2 \in \mathbb{Z}} e(P_\xi(m_1, m_2))\chi_{M'_2}(m_2) = \sum_{r_2=1}^{q} e(P_{a/q}(m_1, r_2)) \sum_{m \in \mathbb{Z}} e(P_{\xi-a/q}(m_1, qm + r_2))\chi_{M'_2}(qm + r_2). \qquad (7.19)$$

The summation in m ranges over $m_* \leq m \leq m_{**}$, where m_*/m_{**} is minimal/maximal with respect to $\tau^{-1}M'_2 \leq qm + r_2 \leq M'_2$. We will use Lemma 2.7 to compare

$$\text{the sum} \quad \sum_{m_* < m \leq m_{**}} e(f(m)) \quad \text{to the integral} \quad \int_{m_*}^{m_{**}} e(f(s))ds,$$

where $f(m) = P_{\xi - a/q}(m_1, qm + r_2)$. Suppose that $a/q \in \Sigma_{\leq l}$ approximates ξ in the following sense:

$$\left| \xi - \frac{a}{q} \right| \leq \frac{(\log_\tau M_2)^\beta}{M_1^{v_{j,1}} M_2^{v_{j,2}}}. \tag{7.20}$$

From the definition of $\Sigma_{\leq l}$, we see that $q \leq 2^{C_\rho 2^{\rho l}}$. Therefore, by (7.20), the derivative f' satisfies

$$|f'(m)| \lesssim q |\xi - a/q| (M_1')^{v_{j,1}} (M_2')^{v_{j,2}-1} \lesssim q (\log_\tau M_2)^\beta M_2^{-1} < 1/2$$

since $v_{j,2} \geq 1$, $\log_2 q \lesssim 2^{\rho l \beta(M_2)} \leq (\log_\tau M_2)^{\rho \beta}$ and $\rho \beta \leq 1/10$ by (7.10). By Lemma 2.7, we have

$$\left| q \sum_{m_* \leq m \leq m_{**}} e(P_{\xi - a/q}(m_1, qm + r_2)) - \int_{\tau^{-1} M_2'}^{M_2'} e(P_{\xi - a/q}(m_1, t)) dt \right| \lesssim q,$$

and hence by (7.19), we obtain

$$\left| \sum_{m_2 \in \mathbb{Z}} e(P_\xi(m_1, m_2)) \chi_{M_2'}(m_2) - G_{m_1}^1(a/q) \mathfrak{m}_{m_1, M_2'}^1(\xi - a/q) \right| \lesssim q(M_2')^{-1},$$

which by $q \leq 2^{C_\rho 2^{\rho l}}$ proves (7.15) as desired. □

For $i \in [2]$ and $j \in [r]$ let $M_1^c = M_2$, $M_2^c = M_1$,

$$\mathbb{S}_\tau^i(j) := \{ M_i \in \mathbb{D}_\tau : (M_1, M_2) \in \mathbb{S}_\tau(j) \text{ for some } M_i^c \}$$

and for $M_1, M_2 \in \mathbb{D}_\tau$ we also let

$$\mathbb{S}_\tau^1(j; M_2) := \{ M_1 \in \mathbb{D}_\tau : (M_1, M_2) \in \mathbb{S}_\tau(j) \},$$
$$\mathbb{S}_\tau^2(j; M_1) := \{ M_2 \in \mathbb{D}_\tau : (M_1, M_2) \in \mathbb{S}_\tau(j) \}.$$

7.4. Changing scale estimates

In our next step, we will have to change the scale (or more precisely, we will truncate the size of denominators of fractions in $\Sigma_{\leq l\beta(M_2)}$) to make the approximation estimates with respect to the first variable possible.

We formulate the change of scale argument as follows.

Claim 7.21. *For every $1 \leq j < r$ and for every $M_1 \in \mathbb{S}_\tau^1(j)$, one has*

$$\| \sup_{M_2 \in \mathbb{S}_\tau^2(j; M_1)} |T_{\mathbb{Z}}[g_{M_1, M_2}^{M_2} - h_{M_1, M_2}^{M_1}] f | \|_{\ell^2(\mathbb{Z})} \lesssim_\tau (\log M_1)^{-\alpha} \|f\|_{\ell^2(\mathbb{Z})}, \qquad f \in \ell^2(\mathbb{Z}), \tag{7.22}$$

with α as in (7.9), where

$$g_{M_1, M_2}^N := m_{M_1, M_2} \Delta_{\leq l\beta(N), \leq -n_{M_1, M_2}^{v_j \cdot \beta}(N)}, \qquad N \geq 1,$$
$$h_{M_1, M_2}^N(\xi) := \sum_{m_1 \in \mathbb{Z}} \Phi_{\leq -n_{M_1, M_2}^{v_j \cdot \beta}(N)}^{\Sigma_{\leq l\beta(N)}} [G_{m_1}^1, \mathfrak{m}_{m_1, M_2}^1](\xi) \chi_{M_1}(m_1), \qquad N \geq 1. \tag{7.23}$$

The same estimate holds when $j = r = 1$, as long as $\log M_1 \leq \log M_2$.

The case $j = r \geq 2$ requires a minor modification. Keeping in mind that $\log M_2 \lesssim \log M_1$, it suffices to establish an analogue of (7.22). Namely, for every $M_2 \in \mathbb{S}_\tau^2(j)$, one has

$$\| \sup_{M_1 \in \mathbb{S}_\tau^1(j;M_2)} |T_\mathbb{Z}[g_{M_1,M_2}^{M_1} - h_{M_1,M_2}^{M_2}]f| \|_{\ell^2(\mathbb{Z})} \lesssim_\tau (\log M_2)^{-\alpha} \|f\|_{\ell^2(\mathbb{Z})}. \tag{7.24}$$

We only present the proof of (7.22); inequality (7.24) can be proved in a similar way.

Proof of Claim 7.21. The proof will proceed in several steps.

Step 1. Using (7.16) from Lemma 7.14, we have

$$\|T_\mathbb{Z}[g_{M_1,M_2}^{M_2} - h_{M_1,M_2}^{M_2}]f\|_{\ell^2(\mathbb{Z})} \lesssim_\tau M_2^{-1/2}\|f\|_{\ell^2(\mathbb{Z})}. \tag{7.25}$$

Hence, by (7.25), it suffices to prove (with α as in (7.9)) that for every $M_1 \in \mathbb{S}_\tau^1(j)$, one has

$$\| \sup_{M_2 \in \mathbb{S}_\tau^2(j;M_1)} |T_\mathbb{Z}[h_{M_1,M_2}^{M_2} - h_{M_1,M_2}^{M_1}]f| \|_{\ell^2(\mathbb{Z})} \lesssim_\tau (\log M_1)^{-\alpha}\|f\|_{\ell^2(\mathbb{Z})}, \qquad f \in \ell^2(\mathbb{Z}). \tag{7.26}$$

Step 2. To prove (7.26), we define for any $s \in \mathbb{N}$ a new multiplier by

$$h_{M_1,M_2,s}^N(\xi) := \sum_{m_1 \in \mathbb{Z}} \Phi_{\leq -n_{M_1,M_2}^{v_j,\beta}(N)}^{\Sigma_s}[G_{m_1}^1, \mathfrak{m}_{m_1,M_2}^1](\xi)\chi_{M_1}(m_1). \tag{7.27}$$

In view of (7.8), we may assume that $l^\beta(M_1) < l^\beta(M_2)$. Then one can write

$$h_{M_1,M_2}^{M_2}(\xi) - h_{M_1,M_2}^{M_1}(\xi) = \sum_{0 \leq s \leq l^\beta(M_1)} \left(h_{M_1,M_2,s}^{M_2}(\xi) - h_{M_1,M_2,s}^{M_1}(\xi)\right) + \sum_{l^\beta(M_1) < s \leq l^\beta(M_2)} h_{M_1,M_2,s}^{M_2}(\xi).$$

For sufficiently large $s \in \mathbb{N}$, if $l^\beta(M_1) \geq s$, then by (7.8) we have $\log_\tau M_2 \geq K_j^{-1}2^{s/\beta} \geq 2^{s/(2\beta)}$. Similarly, if $l^\beta(M_2) \geq s$, then $\log_\tau M_2 \geq 2^{s/(2\beta)}$. Thus, we set $N_s := \tau^{2^{s/(2\beta)}}$ for any $s \in \mathbb{N}$ and let

$$\tilde{\mathbb{S}}_{\tau,M_1}^2(j;s) := \{M_2 \in \mathbb{S}_\tau^2(j;M_1) : M_2 \geq N_s\}.$$

The proof will be finished if we can show (with α and δ as in (7.9)) that for every $f \in \ell^2(\mathbb{Z})$ and for every $M_1 \in \mathbb{S}_\tau^1(j)$, and $0 \leq s \leq l^\beta(M_1)$, one has

$$\| \sup_{M_2 \in \tilde{\mathbb{S}}_{\tau,M_1}^2(j;s)} |T_\mathbb{Z}[h_{M_1,M_2,s}^{M_2} - h_{M_1,M_2,s}^{M_1}]f| \|_{\ell^2(\mathbb{Z})} \lesssim_\tau 2^{-\delta s}(\log M_1)^{-\alpha}\|f\|_{\ell^2(\mathbb{Z})}, \tag{7.28}$$

and moreover, for every $s \in \mathbb{N}$, one also has

$$\| \sup_{M_2 \in \tilde{\mathbb{S}}_{\tau,M_1}^2(j;s)} |T_\mathbb{Z}[h_{M_1,M_2,s}^{M_2}]f| \|_{\ell^2(\mathbb{Z})} \lesssim_\tau s2^{-\delta s}\|f\|_{\ell^2(\mathbb{Z})}. \tag{7.29}$$

Then summing (7.29) over $s \geq l^\beta(M_1)$, we obtain the desired claim by (7.9).

Step 3. We now establish (7.28). If $N \in \{M_1, M_2\}$ and $(M_1, M_2) \in \mathbb{S}_\tau(j)$, then for $M_2 \geq N_s$, we note that

$$\begin{aligned}
\eta_{\leq -n_{M_1,M_2}^{v_j,\beta}(N)}(\xi) &= \eta_{\leq -n_{M_1,M_2}^{v_j,\beta}(N)}(\xi)\eta_{\leq -n_{M_1,M_2}^{v_j,\beta}(N)+1}(\xi) \\
&= \eta_{\leq -n_{M_1,M_2}^{v_j,\beta}(N)}(\xi)\eta_{\leq -n_{M_1,N_s}^{v_j,\beta}(N_s)+1}(\xi),
\end{aligned} \tag{7.30}$$

since $1 \le j < r$ and $v_{j,2} \ne 0$. Using (7.30), we may write

$$
\begin{aligned}
&\Phi^{\Sigma_s}_{\le -n^{v_j,\beta}_{M_1,M_2}(N)}[G^1_{m_1}, \mathfrak{m}^1_{m_1,M_2}](\xi) \\
&= \Phi^{\Sigma_{\le s}}_{\le -n^{v_j,\beta}_{M_1,N_s}(N_s)+2}[1, \mathfrak{m}^1_{m_1,M_2}\eta_{\le -n^{v_j,\beta}_{M_1,M_2}(N)}](\xi) \times \Phi^{\Sigma_s}_{\le -n^{v_j,\beta}_{M_1,N_s}(N_s)+1}[G^1_{m_1}, 1](\xi),
\end{aligned}
\tag{7.31}
$$

for sufficiently large s such that $0 \le s \le l^\beta(M_1)$, which in turn guarantees that $M_2 > N_s$ as we have seen in the previous step. Denote

$$
I(m_1, M_2) := T_{\mathbb{Z}}\left[\Phi^{\Sigma_{\le s}}_{\le -n^{v_j,\beta}_{M_1,N_s}(N_s)+2}\left[1, \sum_{N=l^\beta(M_1)}^{l^\beta(M_2)-1}\mathfrak{m}^1_{m_1,M_2}D_N\left(\eta_{\le -n^{v_j}_{M_1,M_2}(N)}\right)\right]\right],
$$

where (see definitions (7.3) and (7.4))

$$
D_N\left(\eta_{\le -n^{v_j}_{M_1,M_2}(N)}\right) := \eta_{\le -n^{v_j}_{M_1,M_2}(N+1)} - \eta_{\le -n^{v_j}_{M_1,M_2}(N)}.
$$

Using the factorization from (7.31), one sees

$$
\begin{aligned}
&\left\|\sup_{M_2 \in \tilde{\mathbb{S}}^2_{\tau,M_1}(j;s)}|T_{\mathbb{Z}}[h^{M_2}_{M_1,M_2,s} - h^{M_1}_{M_1,M_2,s}]f|\right\|_{\ell^2(\mathbb{Z})} \\
&\le \sum_{m_1 \in \mathbb{Z}}\|I(m_1,M_2)\|_{\ell^2(\mathbb{Z})\to\ell^2(\mathbb{Z};\ell^\infty_{M_2}(\tilde{\mathbb{S}}^2_{\tau,M_1}(j;s)))}\chi_{M_1}(m_1)\left\|T_{\mathbb{Z}}[\Phi^{\Sigma_s}_{\le -n^{v_j,\beta}_{M_1,N_s}(N_s)+1}[G^1_{m_1}, 1]]f\right\|_{\ell^2(\mathbb{Z})}.
\end{aligned}
$$

Using the Ionescu–Wainger multiplier theory (see Theorem 6.5), we conclude that

$$
\sup_{m_1 \in (\tau^{-1}M_1, M_1] \cap \mathbb{Z}}\|I(m_1,M_2)\|_{\ell^2(\mathbb{Z})\to\ell^2(\mathbb{Z};\ell^\infty_{M_2}(\tilde{\mathbb{S}}^2_{\tau,M_1}(j;s)))} \lesssim_\tau (\log M_1)^{-\alpha}
$$

with α as in (7.9), since using standard square function continuous arguments we have

$$
\left\|\left(\sum_{M_2 \in \tilde{\mathbb{S}}^2_{\tau,M_1}(j;s)}\left|T_{\mathbb{R}}\left[\sum_{N=l^\beta(M_1)}^{l^\beta(M_2)-1}\mathfrak{m}^1_{m_1,M_2}D_N\left(\eta_{\le -n^{v_j}_{M_1,M_2}(N)}\right)\right]f\right|^2\right)^{1/2}\right\|_{L^2(\mathbb{R})} \lesssim (\log M_1)^{-\alpha}\|f\|_{L^2(\mathbb{R})}.
$$

Thus, by the Cauchy–Schwarz inequality, Plancherel's theorem and inequality (5.48), we obtain

$$
\begin{aligned}
&\left\|\sup_{M_2 \in \tilde{\mathbb{S}}^2_{\tau,M_1}(j;s)}|T_{\mathbb{Z}}[h^{M_2}_{M_1,M_2,s} - h^{M_1}_{M_1,M_2,s}]f|\right\|_{\ell^2(\mathbb{Z})} \\
&\lesssim (\log M_1)^{-\alpha}\left\|\left(\sum_{m_1 \in \mathbb{Z}}\chi_{M_1}(m_1)|T_{\mathbb{Z}}[\Phi^{\Sigma_s}_{\le -n^{v_j,\beta}_{M_1,N_s}(N_s)+1}[G^1_{m_1}, 1]]f|^2\right)^{1/2}\right\|_{\ell^2(\mathbb{Z})} \\
&\lesssim_\tau 2^{-\delta s}(\log M_1)^{-\alpha}\|f\|_{\ell^2(\mathbb{Z})}
\end{aligned}
$$

with α and δ as in (7.9), which yields (7.28).

Step 4. We now establish (7.29). Using notation from the previous step and denoting

$$
J(m_1, M_2) := T_{\mathbb{Z}}\left[\Phi^{\Sigma_{\le s}}_{\le -n^{v_j,\beta}_{M_1,N_s}(N_s)+2}\left[1, \mathfrak{m}^1_{m_1,M_2}\eta_{\le n^{v_j,\beta}_{M_1,M_2}(M_2)}\right]\right],
$$

and again using the factorization from (7.31), one sees

$$
\begin{aligned}
\Big\| \sup_{M_2 \in \tilde{\mathbb{S}}^2_{\tau,M_1}(j;s)} |T_{\mathbb{Z}}[h^{M_2}_{M_1,M_2,s}]f| \Big\|_{\ell^2(\mathbb{Z})} &\leq \sum_{m_1 \in \mathbb{Z}} \|J(m_1, M_2)\|_{\ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z};\ell^\infty_{M_2}(\tilde{\mathbb{S}}^2_{\tau,M_1}(j;s)))} \chi_{M_1}(m_1) \\
&\quad \times \Big\| T_{\mathbb{Z}}\big[\Phi^{\Sigma_s}_{\leq -n^{v_j,\beta}_{M_1,N_s}(N_s)+1}[G^1_{m_1},1]\big]f \Big\|_{\ell^2(\mathbb{Z})}.
\end{aligned}
$$

Using the Ionescu–Wainger multiplier theory (see Theorem 6.14), we conclude that

$$
\sup_{m_1 \in (\tau^{-1}M_1, M_1] \cap \mathbb{Z}} \|J(m_1, M_2)\|_{\ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z};\ell^\infty_{M_2}(\tilde{\mathbb{S}}^2_{\tau,M_1}(j;s)))} \lesssim_\tau s.
$$

Then proceeding as in the previous step, we obtain (7.29). This completes the proof of Claim 7.21.   □

7.5. Transition estimates

Our aim will be to understand the final approximation, which will allow us to apply the oscillation Ionescu–Wainger theory (see Theorem 6.37) from Section 6.

Claim 7.32. *For every $1 \leq j < r$ and for every $M_1 \in \mathbb{S}^1_\tau(j)$, one has*

$$
\Big\| \sup_{M_2 \in \mathbb{S}^2_\tau(j;M_1)} |T_{\mathbb{Z}}[h^{M_1}_{M_1,M_2} - \tilde{h}^{M_1}_{M_1,M_2}]f| \Big\|_{\ell^2(\mathbb{Z})} \lesssim_\tau (\log M_1)^{-\alpha} \|f\|_{\ell^2(\mathbb{Z})}, \qquad f \in \ell^2(\mathbb{Z}), \tag{7.33}
$$

with α as in (7.9), where $h^N_{M_1,M_2}$ was defined in (7.23) and

$$
\tilde{h}^N_{M_1,M_2} := \Phi^{\Sigma_{\leq l^\beta(N)}}_{\leq -n^{v_j,\beta}_{M_1,M_2}(N)}[G, \mathfrak{m}_{M_1,M_2}], \qquad N \geq 1. \tag{7.34}
$$

The same estimate holds when $j = r = 1$, as long as $\log M_1 \leq \log M_2$.

The case $j = r \geq 2$ requires a minor modification. Keeping in mind that $\log M_2 \lesssim \log M_1$, it suffices to establish an analogue of (7.33). Namely, one has

$$
\Big\| \sup_{M_1 \in \mathbb{S}^1_\tau(j;M_2)} |T_{\mathbb{Z}}[h^{M_2}_{M_1,M_2} - \tilde{h}^{M_2}_{M_1,M_2}]f| \Big\|_{\ell^2(\mathbb{Z})} \lesssim_\tau (\log M_2)^{-\alpha} \|f\|_{\ell^2(\mathbb{Z})}, \qquad f \in \ell^2(\mathbb{Z}). \tag{7.35}
$$

We only present the proof of (7.33); inequality (7.35) can be proved in a similar way.

Proof of Claim 7.32. The proof will proceed in several steps as before. Write

$$
h^{M_1}_{M_1,M_2} - \tilde{h}^{M_1}_{M_1,M_2} = \sum_{0 \leq s \leq l^\beta(M_1)} h^{M_1}_{M_1,M_2,s} - \tilde{h}^{M_1}_{M_1,M_2,s},
$$

where $h^{M_1}_{M_1,M_2,s}$ was defined in (7.27) and

$$
\tilde{h}^{M_1}_{M_1,M_2,s} := \Phi^{\Sigma_s}_{\leq -n^{v_j,\beta}_{M_1,M_2}(M_1)}[G, \mathfrak{m}_{M_1,M_2}]. \tag{7.36}
$$

Then it suffices to show that for sufficiently large s such that $0 \leq s \leq l^\beta(M_1)$, we have

$$
\Big\| \sup_{M_2 \in \mathbb{S}^2_\tau(j;M_1)} |T_{\mathbb{Z}}[h^{M_1}_{M_1,M_2,s} - \tilde{h}^{M_1}_{M_1,M_2,s}]f| \Big\|_{\ell^2(\mathbb{Z})} \lesssim_\tau s2^{-\delta s}(\log M_1)^{-\alpha}\|f\|_{\ell^2(\mathbb{Z})}, \qquad f \in \ell^2(\mathbb{Z}),
$$

$$
\tag{7.37}
$$

with α and δ as in (7.9), which will clearly imply (7.33).

Step 1. Using (7.30), in a similar way as in (7.31), we may write

$$\tilde{h}^{M_1}_{M_1,M_2,s}(\xi) = \Phi^{\Sigma_{\leq s}}_{\leq -n^{v_j,\beta}_{M_1,N_s}(N_s)+2}[1, \widetilde{\mathfrak{m}}_{M_1,M_2}](\xi) \times \Phi^{\Sigma_s}_{\leq -n^{v_j,\beta}_{M_1,N_s}(N_s)+1}[G,1](\xi), \tag{7.38}$$

where

$$\widetilde{\mathfrak{m}}_{M_1,M_2}(\xi) := \mathfrak{m}_{M_1,M_2}(\xi)\eta_{\leq -n^{v_j,\beta}_{M_1,M_2}(M_1)}(\xi). \tag{7.39}$$

By Theorem 6.14, we may conclude

$$\left\| T_{\mathbb{Z}}\left[\Phi^{\Sigma_{\leq s}}_{\leq -n^{v_j,\beta}_{M_1,N_s}(N_s)+2}[1, \widetilde{\mathfrak{m}}_{M_1,M_2}] \right] \right\|_{\ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z}); \ell^\infty_{M_2}(\mathbb{S}^2_\tau(j;M_1))} \lesssim_\tau s. \tag{7.40}$$

By Plancherel's theorem and inequality (5.47), we obtain

$$\left\| T_{\mathbb{Z}}\left[\Phi^{\Sigma_s}_{\leq -n^{v_j,\beta}_{M_1,N_s}(N_s)+1}[G,1] \right] f \right\|_{\ell^2(\mathbb{Z})} \lesssim_\tau 2^{-\delta s} \|f\|_{\ell^2(\mathbb{Z})}. \tag{7.41}$$

Inequalities (7.40) and (7.41) and (7.38) imply

$$\left\| \sup_{M_2 \in \mathbb{S}^2_\tau(j;M_1)} \left| T_{\mathbb{Z}}[\tilde{h}^{M_1}_{M_1,M_2,s}] f \right| \right\|_{\ell^2(\mathbb{Z})} \lesssim_\tau s2^{-\delta s} \|f\|_{\ell^2(\mathbb{Z})}, \qquad f \in \ell^2(\mathbb{Z}). \tag{7.42}$$

Step 2. We now establish (7.37). For $0 \leq s \leq l^\beta(M_1)$, we note that

$$h^{M_1}_{M_1,M_2,s}(\xi) = \sum_{a/q \in \Sigma_s} \eta_{\leq -n^{v_j,\beta}_{M_1,M_2}(M_1)}(\xi - a/q) \sum_{r_1=1}^q G^1_{r_1}(a/q)$$
$$\times \sum_{m_1 \in \mathbb{Z}} \mathfrak{m}^1_{qm_1+r_1,M_2}(\xi - a/q)\chi_{M_1}(qm_1 + r_1).$$

Introducing $\theta := \xi - a/q$, $U_1 := \frac{\tau^{-1}M_1 - r_1}{q}$ and $V_1 := \frac{M_1 - r_1}{q}$, one can expand

$$\mathfrak{m}^1_{qm_1+r_1,M_2}(\theta) = \frac{1}{1 - \tau^{-1}} \int_{\tau^{-1}}^1 e(P_\theta(qm_1 + r_1, M_2 y_2))dy_2,$$

and by the fundamental theorem of calculus, one can write

$$\sum_{\lfloor U_1 \rfloor < m_1 \leq \lfloor V_1 \rfloor} \int_{\tau^{-1}}^1 e(P_\theta(qm_1 + r_1, M_2 y_2))dy_2 - \int_{U_1}^{V_1} \int_{\tau^{-1}}^1 e(P_\theta(qy_1 + r_1, M_2 y_2))dy_2 dy_1$$

$$= \sum_{\lfloor U_1 \rfloor < m_1 \leq \lfloor V_1 \rfloor} \int_{m_1-1}^{m_1} \int_{y_1}^{m_1} \int_{\tau^{-1}}^1 2\pi i q\theta(\partial_1 P)(qt + r_1, M_2 y_2)e(P_\theta(qt + r_1, M_2 y_2))dy_2 dt dy_1$$

$$+ \left(\int_{\lfloor U_1 \rfloor}^{U_1} - \int_{\lfloor V_1 \rfloor}^{V_1} \right) \int_{\tau^{-1}}^1 e(P_\theta(qy_1 + r_1, M_2 y_2))dy_2 dy_1.$$

By the change of variable, we have

$$\int_{U_1}^{V_1} \int_{\tau^{-1}}^1 e(P_\theta(qy_1 + r_1, M_2 y_2))dy_2 dy_1 = \frac{M_1(1 - \tau^{-1})^2}{q}\mathfrak{m}_{M_1,M_2}(\theta).$$

We now define new multipliers

$$\mathfrak{g}_{M_1,M_2}^{r_1,1}(\theta)$$

$$:= \sum_{\lfloor U_1 \rfloor < m_1 \leq \lfloor V_1 \rfloor} \int_{m_1-1}^{m_1} \int_{y_1}^{m_1} \int_{\tau^{-1}}^{1} \frac{2\pi i q (\log M_1)^\beta (\partial_1 P)(qt + r_1, M_2 y_2) e(P_\theta(qt + r_1, M_2 y_2))}{(1 - \tau^{-1}) M_1^{v_{j,1}} M_2^{v_{j,2}} |(\tau^{-1}M_1, M_1] \cap \mathbb{Z}|} \, dy_2 dt dy_1,$$

and finally

$$\mathfrak{g}_{M_1,M_2}^{r_1,2}(\theta) := \Big(\int_{\lfloor U_1 \rfloor}^{U_1} - \int_{\lfloor V_1 \rfloor}^{V_1} \Big) \int_{\tau^{-1}}^{1} \frac{e(P_\theta(qy_1 + r_1, M_2 y_2))}{(1 - \tau^{-1})|(\tau^{-1}M_1, M_1] \cap \mathbb{Z}|} \, dy_2 dy_1.$$

Then with these definitions, we can write

$$h_{M_1,M_2,s}^{M_1}(\xi) - \tilde{h}_{M_1,M_2,s}^{M_1}(\xi) = \gamma_{\tau,M_1} \tilde{h}_{M_1,M_2,s}^{M_1}(\xi) + \sum_{\ell \in [2]} \sum_{a/q \in \Sigma_s} \sum_{r_1=1}^{q} G_{r_1}^1(a/q) \mathfrak{h}_{M_1,M_2}^{r_1,\ell}(\xi - a/q),$$

where $\gamma_{\tau,M_1} := \frac{\{M_1\} - \{\tau^{-1}M_1\}}{|(\tau^{-1}M_1, M_1] \cap \mathbb{Z}|}$ and

$$\mathfrak{h}_{M_1,M_2}^{r_1,1}(\theta) := \mathfrak{g}_{M_1,M_2}^{r_1,1}(\theta) \varrho_{\leq -n_{M_1,M_2}^{v_j,\beta}(M_1)}(\theta) \quad \text{and} \quad \mathfrak{h}_{M_1,M_2}^{r_1,2}(\theta) := \mathfrak{g}_{M_1,M_2}^{r_1,2}(\theta) \eta_{\leq -n_{M_1,M_2}^{v_j,\beta}(M_1)}(\theta)$$

and $\varrho_{\leq n}(\theta) := (2^{-n}\theta)\eta_{\leq n}(\theta)$. For $\ell \in [2]$, we have

$$|\mathfrak{h}_{M_1,M_2}^{r_1,\ell}(\theta)| \lesssim q(\log M_1)^\beta M_1^{-1} \quad \text{and} \quad |\gamma_{\tau,M_1}| \lesssim_\tau M_1^{-1}.$$

Finally, using Theorem 6.14 for each $\ell \in [2]$, we conclude

$$\Big\| \sup_{M_2 \in \mathbb{S}_\tau^2(j;M_1)} \Big| T_\mathbb{Z} \Big[\sum_{\ell \in [2]} \sum_{a/q \in \Sigma_s} \sum_{r_1=1}^{q} G_{r_1}^1(a/q) \mathfrak{h}_{M_1,M_2}^{r_1,\ell}(\cdot - a/q) \Big] f \Big| \Big\|_{\ell^2(\mathbb{Z})} \lesssim 2^{-\delta s} M_1^{-3/4} \|f\|_{\ell^2(\mathbb{Z})}.$$

This in turn, combined with (7.42), implies (7.37), and the proof of Claim 7.32 is established. □

7.6. All together: Proof of Theorem 7.6

We begin with a useful auxiliary lemma.

Lemma 7.43. *For every $p \in (1, \infty)$ and every $j \in [r]$, there exists a constant $\delta_P \in (0, 1)$ such that for every $f \in \ell^p(\mathbb{Z})$ and $s \in \mathbb{N}$, one has*

$$\Big\| T_\mathbb{Z} \Big[\Phi_{\leq -n_{N_s,N_s}^{v_j,\beta}(N_s)+1}^{\Sigma_s} [G, \Pi_s^\beta] \Big] f \Big\|_{\ell^p(\mathbb{Z})} \lesssim_{p,\tau} 2^{-\delta_P s} \|f\|_{\ell^p(\mathbb{Z})}, \tag{7.44}$$

where $N_s := \tau^{2^{s/(2\beta)}}$ for any $s \in \mathbb{N}$, and $\Pi_s^\beta(\xi) := \prod_{u \in S_P} \eta_{\leq -n_{N_s,N_s}^{u,\beta}(N_s)+1}(\xi)$ with $\beta > 0$ from (7.9).

Proof. We may assume that $s \geq 0$ is large; otherwise, there is nothing to prove. Inequality (7.44) for $p = 2$ with $\delta_2 = \delta$ as in Proposition 5.46 follows by Plancherel's theorem from inequality (5.47) and the disjointness of supports of $\Pi_s^\beta(\xi - a/q)$ whenever $a/q \in \Sigma_s$.

We now prove (7.44) for $p \neq 2$. We shall proceed in four steps.

Step 1. Let $M \simeq 2^{10 C_\rho 2^{10\rho s}}$ define

$$\mathfrak{h}_M^s := m_{M,M} \Phi_{\leq -n_{N_s,N_s}^{v_j,\beta}(N_s)+1}^{\Sigma_s} [1, \Pi_s^\beta].$$

By the Ionescu–Wainger multiplier theorem (see Theorem 6.5), one has

$$\left\| T_{\mathbb{Z}}[\mathfrak{h}^s_M] f \right\|_{\ell^u(\mathbb{Z})} \lesssim_{u,\tau} \|f\|_{\ell^u(\mathbb{Z})}, \tag{7.45}$$

whenever $u \in \{p_0, p'_0\}$. We will prove

$$\left\| T_{\mathbb{Z}}\big[\mathfrak{h}^s_M - \Phi^{\Sigma_s}_{\leq -n^{v_j,\beta}_{N_s,N_s}(N_s)+1}[G, \mathfrak{m}_{M,M}\Pi^\beta_s]\big] f \right\|_{\ell^p(\mathbb{Z})} \lesssim_{p,\tau} \|f\|_{\ell^p(\mathbb{Z})} \tag{7.46}$$

and

$$\left\| T_{\mathbb{Z}}\big[\Phi^{\Sigma_s}_{\leq -n^{v_j,\beta}_{N_s,N_s}(N_s)+1}[G, (1-\mathfrak{m}_{M,M})\Pi^\beta_s]\big] f \right\|_{\ell^p(\mathbb{Z})} \lesssim_{p,\tau} \|f\|_{\ell^p(\mathbb{Z})}. \tag{7.47}$$

Assuming momentarily that (7.46) and (7.47) hold, then (7.45) and the triangle inequality yield

$$\left\| T_{\mathbb{Z}}\big[\Phi^{\Sigma_s}_{\leq -n^{v_j,\beta}_{N_s,N_s}(N_s)+1}[G, \Pi^\beta_s]\big] f \right\|_{\ell^u(\mathbb{Z})} \lesssim_{u,\tau} \|f\|_{\ell^u(\mathbb{Z})}, \tag{7.48}$$

whenever $u \in \{p_0, p'_0\}$. Then interpolation between (7.44) for $p = 2$ (that we have shown with $\delta_2 = \delta$) and (7.48) gives (7.44) for all $p \in (1, \infty)$.

Step 2. We now establish (7.46). For $p = 2$, it will suffice to show that

$$\left| m_{M,M}(\xi)\Phi^{\Sigma_s}_{\leq -n^{v_j,\beta}_{N_s,N_s}(N_s)+1}[1, \Pi^\beta_s](\xi) - \Phi^{\Sigma_s}_{\leq -n^{v_j,\beta}_{N_s,N_s}(N_s)+1}[G, \mathfrak{m}_{M,M}\Pi^\beta_s](\xi) \right| \lesssim 2^{-5C_\rho 2^{5\rho s}}. \tag{7.49}$$

Then by (7.49) and Plancherel's theorem, we obtain for sufficiently large $s \in \mathbb{N}$ that

$$\left\| T_{\mathbb{Z}}\big[\mathfrak{h}^s_M - \Phi^{\Sigma_s}_{\leq -n^{v_j,\beta}_{N_s,N_s}(N_s)+1}[G, \mathfrak{m}_{M,M}\Pi^\beta_s]\big] f \right\|_{\ell^2(\mathbb{Z})} \lesssim_\tau 2^{-5C_\rho 2^{5\rho s}} \|f\|_{\ell^2(\mathbb{Z})}. \tag{7.50}$$

Moreover, for $u \in \{p_0, p'_0\}$, we have the trivial estimate

$$\left\| T_{\mathbb{Z}}\big[\mathfrak{h}^s_M - \Phi^{\Sigma_s}_{\leq -n^{v_j,\beta}_{N_s,N_s}(N_s)+1}[G, \mathfrak{m}_{M,M}\Pi^\beta_s]\big] f \right\|_{\ell^u(\mathbb{Z})} \lesssim_{u,\tau} 2^{2C_\rho 2^{\rho s}} \|f\|_{\ell^u(\mathbb{Z})}, \tag{7.51}$$

due to (6.4). Interpolating (7.50) and (7.51) gives (7.46).

Step 3. To prove (7.49), we proceed as in the proof of Lemma 7.14 and show that

$$\left| m_{M,M}(\xi) - G(a/q)\mathfrak{m}_{M,M}(\xi - a/q) \right| \lesssim qM^{-1}, \tag{7.52}$$

whenever $a/q \in \Sigma_s$ and $|\xi - a/q| \leq \min_{u \in S_P}\{(\log_\tau N_s)^\beta N_s^{-u_1} N_s^{-u_2}\}$. Then (7.52) immediately gives (7.49), since $q \leq 2^{C_\rho 2^{\rho s}}$ if $a/q \in \Sigma_s$. To verify (7.52), we use Lemma 2.7 twice, which can be applied, since the derivatives $\partial_{m_1} f$ and $\partial_{m_2} f$ of $f(m_1, m_2) = P_{\xi-a/q}(qm_1 + r_1, qm_2 + r_2)$ satisfy

$$\left| \partial_{m_\ell} f(m_1, m_2) \right| \lesssim q|\xi - a/q| \sum_{u \in S_P} M^{u_1+u_2-1} \lesssim q(\log_\tau N_s)^\beta N_s^{-1} < 1/2, \qquad \ell \in [2]$$

for sufficiently large $s \in \mathbb{N}$, since $M \leq N_s^{1/5}$, $q \leq 2^{C_\rho 2^{\rho s}}$ and $\rho\beta \leq 1/10$ by (7.10), and we are done.

Step 4. We now establish (7.47). Assume that $p = 2$ and observe that

$$\left| (1-\mathfrak{m}_{M,M}(\xi - a/q))\Pi^\beta_s(\xi - a/q) \right| \lesssim |\xi - a/q| \sum_{u \in S_P} M^{u_1+u_2} \lesssim N_s^{-3/4} \lesssim 2^{-10C_\rho 2^{5\rho s}}$$

for sufficiently large $s \in \mathbb{N}$, since $M \simeq 2^{10C_\rho 2^{10\rho s}}$, and $\rho\beta < 1/1000$. Using this bound and Plancherel's theorem, we see that

$$\left\| T_{\mathbb{Z}} \left[\Phi^{\Sigma_s}_{\leq -n^{v_j,\beta}_{N_s,N_s}(N_s)+1} [G, (1 - \mathfrak{m}_{M,M})\Pi^\beta_s] \right] f \right\|_{\ell^2(\mathbb{Z})} \lesssim_\tau 2^{-5C_\rho 2^{5\rho s}} \|f\|_{\ell^2(\mathbb{Z})}. \tag{7.53}$$

Moreover, by (6.4), for $u \in \{p_0, p'_0\}$, we have the trivial estimate

$$\left\| T_{\mathbb{Z}} \left[\Phi^{\Sigma_s}_{\leq -n^{v_j,\beta}_{N_s,N_s}(N_s)+1} [G, (1 - \mathfrak{m}_{M,M})\Pi^\beta_s] \right] f \right\|_{\ell^u(\mathbb{Z})} \lesssim_{u,\tau} 2^{2C_\rho 2^{\rho s}} \|f\|_{\ell^u(\mathbb{Z})}. \tag{7.54}$$

Interpolation between (7.53) and (7.54) yields (7.47), and the proof of Lemma 7.43 is complete. □

Recalling the definition of $\tilde{h}^{M_1}_{M_1,M_2}$ from (7.34), we now prove the following claim:

Claim 7.55. *For every $p \in (1, \infty)$ and every $1 \leq j < r$ and for every $f \in \ell^p(\mathbb{Z})$, one has*

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{S}_\tau(j))} \| O_{I,J}(T_{\mathbb{Z}}[\tilde{h}^{M_1}_{M_1,M_2}]f : (M_1, M_2) \in \mathbb{S}_\tau(j)) \|_{\ell^p(\mathbb{Z})} \lesssim_{p,\tau} \|f\|_{\ell^p(\mathbb{Z})}. \tag{7.56}$$

The same estimate holds when $j = r = 1$, as long as $\log M_1 \leq \log M_2$.

When $j = r \geq 2$, in view of (7.35), we will be able to reduce the problem to the following:

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\mathbb{S}_\tau(j))} \| O_{I,J}(T_{\mathbb{Z}}[\tilde{h}^{M_2}_{M_1,M_2}]f : (M_1, M_2) \in \mathbb{S}_\tau(j)) \|_{\ell^p(\mathbb{Z})} \lesssim_{p,\tau} \|f\|_{\ell^p(\mathbb{Z})}. \tag{7.57}$$

We will only prove (7.56); the proof of (7.57) will follow in a similar way. We omit details.

Proof of Claim 7.55. The proof will consist of two steps to make the argument clear.

Step 1. Similarly as in Claim 7.21, we define $N_s := \tau^{2^{s/(2\beta)}}$ for any $s \in \mathbb{N}$ and introduce

$$\tilde{\mathbb{S}}_\tau(j, s) := \{(M_1, M_2) \in \mathbb{S}_\tau(j) : M_1 \geq N_s\}.$$

For each $(M_1, M_2) \in \mathbb{S}_\tau(j)$, we have $M_1^{v_{j,1}} M_2^{v_{j,2}} \geq M_1^{u_1} M_2^{u_2}$ for every $u = (u_1, u_2) \in S_P$. Hence,

$$\begin{aligned}
\eta_{\leq -n^{v_j,\beta}_{M_1,M_2}(M_1)}(\xi) &= \eta_{\leq -n^{v_j,\beta}_{M_1,M_2}(M_1)}(\xi) \prod_{u \in S_P} \eta_{\leq -n^{u,\beta}_{M_1,M_2}(M_1)+1}(\xi) \\
&= \eta_{\leq -n^{v_j,\beta}_{M_1,M_2}(M_1)}(\xi)\Pi^\beta_s(\xi)
\end{aligned} \tag{7.58}$$

holds for sufficiently large $s \in \mathbb{N}$ so that $0 \leq s \leq l^\beta(M_1)$, where Π^β_s was defined in Lemma 7.43.

The proof of (7.56) will be completed if we show (with $\tilde{h}^{M_1}_{M_1,M_2,s}$ defined in (7.36)) that for every $p \in (1, \infty)$, there is $\delta_p \in (0, 1)$ such that for all $f \in \ell^p(\mathbb{Z})$, we have

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\tilde{\mathbb{S}}_\tau(j,s))} \| O_{I,J}(T_{\mathbb{Z}}[\tilde{h}^{M_1}_{M_1,M_2,s}]f : (M_1, M_2) \in \tilde{\mathbb{S}}_\tau(j, s)) \|_{\ell^p(\mathbb{Z})} \lesssim_{p,\tau} s2^{-\delta_p s} \|f\|_{\ell^p(\mathbb{Z})}. \tag{7.59}$$

Using $\widetilde{\mathfrak{m}}_{M_1,M_2}$ from (7.39) and (7.58), we may write

$$\tilde{h}^{M_1}_{M_1,M_2,s}(\xi) = \Phi^{\Sigma_{\leq s}}_{\leq -n^{v_j,\beta}_{N_s,N_s}(N_s)+2} [1, \widetilde{\mathfrak{m}}_{M_1,M_2}](\xi) \times \Phi^{\Sigma_s}_{\leq -n^{v_j,\beta}_{N_s,N_s}(N_s)+1} [G, \Pi^\beta_s](\xi). \tag{7.60}$$

By Lemma 7.43, for sufficiently large $s \in \mathbb{N}$, we have

$$\left\| T_{\mathbb{Z}} \left[\Phi^{\Sigma_s}_{\leq -n^{v_j,\beta}_{N_s,N_s}(N_s)+1} [G, \Pi^\beta_s] \right] f \right\|_{\ell^p(\mathbb{Z})} \lesssim_{p,\tau} 2^{-\delta_p s} \|f\|_{\ell^p(\mathbb{Z})}. \tag{7.61}$$

Using factorization (7.60) and (7.61), it suffices to prove that

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\tilde{\mathbb{S}}_\tau(j,s))} \|O_{I,J}(T_\mathbb{Z}[\Phi^{\Sigma_{\leq s}}_{\leq -n^{v_j,\beta}_{N_s,N_s}(N_s)+2}[1, \widetilde{\mathfrak{m}}_{M_1,M_2}]]f : (M_1, M_2) \in \tilde{\mathbb{S}}_\tau(j,s))\|_{\ell^P(\mathbb{Z})}$$

$$\lesssim_{p,\tau} s\|f\|_{\ell^P(\mathbb{Z})},$$

which will readily imply (7.59).

Step 2. Appealing to the Ionescu–Wainger multiplier theory (see Theorem 6.37) for oscillation seminorms developed in the previous section, we see that

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\tilde{\mathbb{S}}_\tau(j,s))} \|O_{I,J}(T_\mathbb{Z}[\Phi^{\Sigma_{\leq s}}_{\leq -n^{v_j,\beta}_{N_s,N_s}(N_s)+2}[1, \mathfrak{m}_{M_1,M_2}]]f : (M_1, M_2) \in \tilde{\mathbb{S}}_\tau(j,s))\|_{\ell^P(\mathbb{Z})}$$

$$\lesssim_{p,\tau} s\|f\|_{\ell^P(\mathbb{Z})}.$$

Hence, the last inequality from the previous step will be proved if we establish

$$\sup_{J \in \mathbb{Z}_+} \sup_{I \in \mathfrak{S}_J(\tilde{\mathbb{S}}_\tau(j,s))} \|O_{I,J}(T_\mathbb{Z}[\Phi^{\Sigma_{\leq s}}_{\leq -n^{v_j,\beta}_{N_s,N_s}(N_s)+2}[1, \mathfrak{g}_{M_1,M_2}]]f : (M_1, M_2) \in \tilde{\mathbb{S}}_\tau(j,s))\|_{\ell^P(\mathbb{Z})}$$

$$\lesssim_{p,\tau} \|f\|_{\ell^P(\mathbb{Z})}, \tag{7.62}$$

with $\mathfrak{g}_{M_1,M_2} = \widetilde{\mathfrak{m}}_{M_1,M_2} - \mathfrak{m}_{M_1,M_2}$. By the van der Corput estimate (Proposition 2.6) for $\mathfrak{m}_{M_1,M_2}$, there exists $\delta_0 > 0$ (in fact, $\delta_0 \simeq (\deg P)^{-1}$) such that

$$|\mathfrak{g}_{M_1,M_2}(\xi)| = |\mathfrak{m}_{M_1,M_2}(\xi)(1 - \eta_{\leq -n^{v_j,\beta}_{M_1,M_2}(M_1)}(\xi))| \lesssim \min\{(\log M_1)^{-\delta_0\beta}, (M_1^{v_{j,1}} M_2^{v_{j,2}} |\xi|)^{\pm\delta_0}\}$$

for $(M_1, M_2) \in \tilde{\mathbb{S}}_\tau(j)$, since

$$|1 - \eta_{\leq -n^{v_j,\beta}_{M_1,M_2}(M_1)}(\xi)| \lesssim \min\{1, M_1^{v_{j,1}} M_2^{v_{j,2}} |\xi|\}.$$

Then by Plancherel's theorem combined with a simple interpolation and Theorem 6.5, we conclude that for every $p \in (1, \infty)$, there is $\alpha_p > 10$ such that for every $f \in \ell^p(\mathbb{Z})$, one has

$$\left\| \left(\sum_{M_2 \in \tilde{\mathbb{S}}^2_\tau(j;M_1)} |T_\mathbb{Z}[\Phi^{\Sigma_{\leq s}}_{\leq -n^{v_j,\beta}_{N_s,N_s}(N_s)+2}[1, \mathfrak{g}_{M_1,M_2}]]f|^2 \right)^{1/2} \right\|_{\ell^P(\mathbb{Z})} \lesssim_{p,\tau} (\log M_1)^{-\alpha_p} \|f\|_{\ell^P(\mathbb{Z})},$$

completing the proof of (7.62). □

Proof of Theorem 7.6. We fix $1 \leq j < r$ as before. To prove (7.7), in view of (7.56) and (2.12), it suffices to show that

$$\sum_{M_1 \in \mathbb{S}^1_\tau(j)} \left\| \sup_{M_2 \in \mathbb{S}^2_\tau(j;M_1)} |T_\mathbb{Z}[m_{M_1,M_2} - \tilde{h}^{M_1}_{M_1,M_2}]f| \right\|_{\ell^P(\mathbb{Z})} \lesssim_{p,\tau} \|f\|_{\ell^P(\mathbb{Z})}. \tag{7.63}$$

For $u \in \{p_0, p'_0\}$ by the one-parameter theory, which produces bounds independent of the coefficients of the underlying polynomials (see, for instance, [52, 47]), we may conclude

$$\sup_{M_1 \in \mathbb{Z}_+} \left\| \sup_{M_2 \in \mathbb{Z}_+} |T_\mathbb{Z}[m_{M_1,M_2}]f| \right\|_{\ell^u(\mathbb{Z})} \lesssim_{u,\tau} \|f\|_{\ell^u(\mathbb{Z})}, \tag{7.64}$$

and by (2.17) combined with (7.56), we also have

$$\sup_{M_1 \in \mathbb{S}^1_\tau(j)} \left\| \sup_{M_2 \in \mathbb{S}^2_\tau(j;M_1)} |T_\mathbb{Z}[\tilde{h}^{M_1}_{M_1,M_2}]f| \right\|_{\ell^u(\mathbb{Z})} \lesssim_{u,\tau} \|f\|_{\ell^u(\mathbb{Z})}. \tag{7.65}$$

On the one hand, combining (7.64) and (7.65), we deduce that

$$\Big\| \sup_{M_2 \in \mathbb{S}^2_\tau(j;M_1)} |T_{\mathbb{Z}}[m_{M_1,M_2} - \tilde{h}^{M_1}_{M_1,M_2}]f| \Big\|_{\ell^u(\mathbb{Z})} \lesssim_{u,\tau} \|f\|_{\ell^u(\mathbb{Z})}. \tag{7.66}$$

On the other hand, inequalities (7.12), (7.22) and (7.33) imply for every $M_1 \in \mathbb{S}^1_\tau(j)$ that

$$\Big\| \sup_{M_2 \in \mathbb{S}^2_\tau(j;M_1)} |T_{\mathbb{Z}}[m_{M_1,M_2} - \tilde{h}^{M_1}_{M_1,M_2}]f| \Big\|_{\ell^2(\mathbb{Z})} \lesssim_\tau (\log M_1)^{-\alpha} \|f\|_{\ell^2(\mathbb{Z})} \tag{7.67}$$

with the parameter $\alpha > 0$ as in (7.9). Simple interpolation between (7.66) and (7.67) yields (7.63), and this completes the proof of Theorem 7.6. □

References

[1] G. I. Arkhipov, V. N. Chubarikov and A. A. Karatsuba, *Trigonometric Sums in Number Theory and Analysis (De Gruyter Expositions in Mathematics)* (Walter De Gruyter, 2004).

[2] G. I. Arkhipov, V. N. Chubarikov and A. A. Karatsuba, 'Distribution of fractional parts of polynomials of several variable', *Mat. Zametki* **25**(1) (1979), 3–14

[3] T. Austin, 'A proof of Walsh's convergence theorem using couplings', *Int. Math. Res. Not. IMRN* **15** (2015), 6661–6674.

[4] T. Austin, 'On the norm convergence of non-conventional ergodic averages', *Ergodic Theory Dynam. Systems* **30** (2010), 321–338.

[5] A. Bellow, *Measure Theory: Proceedings of the Conference Held at Oberwolfach, June 21–27, 1981* (Lecture Notes in Mathematics) vol. **541** (Springer-Verlag, Berlin, 1982). Section: Two problems submitted by A. Bellow, 429–431.

[6] V. Bergelson, 'Weakly mixing PET', *Ergodic Theory Dynam. Systems* **7**(3) (1987), 337–349.

[7] V. Bergelson, 'Ergodic Ramsey Theory – An Update', in M. Pollicott and K. Schmidt (eds.) *Ergodic Theory of $\mathbb{Z}^d$-actions* (London Math. Soc. Lecture Note Series) vol. **228** (1996), 1–61.

[8] V. Bergelson, 'Combinatorial and diophantine applications of ergodic theory', in B. Hasselblatt and A. Katok (eds.) *Handbook of Dynamical Systems* vol. **1B** (Elsevier, 2006), 745–841.

[9] V. Bergelson and A. Leibman, 'Polynomial extensions of van derWaerden's and Szemerédi's theorems', *J. Amer. Math. Soc.* **9** (1996), 725–753.

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

[11] G. Birkhoff, 'Proof of the ergodic theorem', *Proc. Natl. Acad. Sci. USA* **17**(12) (1931), 656–660.

[12] P. Bohl, 'Über ein in der Theorie der säkularen Störungen vorkommendes Problem', *J. Reine Angew. Math.* **135** (1909), 189–283.

[13] J. Bourgain, 'On the maximal ergodic theorem for certain subsets of the integers', *Israel J. Math.* **61** (1988), 39–72.

[14] J. Bourgain, 'On the pointwise ergodic theorem on L^p for arithmetic sets', *Israel J. Math.* **61** (1988), 73–84.

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

[16] J. Bourgain, 'Double recurrence and almost sure convergence', *J. Reine Angew. Math.* **404** (1990), 140–161.

[17] J. Bourgain, C. Demeter and L. Guth, 'Proof of the main conjecture in Vinogradov's mean value theorem for degrees higher than three', *Ann. Math.* **184**(2) (2016), 633–682.

[18] Z. Buczolich and R. D. Mauldin, 'Divergent square averages', *Ann. Math.* **171**(3) (2010), 1479–1530.

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

[20] A. Carbery, M. Christ and J. Wright, 'Multidimensional van der Corput and sublevel set estimates', *J. Amer. Math. Soc.* **12**(4) (1999), 3981–1015.

[21] Q. Chu, N. Frantzikinakis and B. Host, 'Ergodic averages of commuting transformations with distinct degree polynomial iterates', *Proc. London. Math. Soc.* **102**(5) (2011), 801–842.

[22] J. Duoandikoetxea and J. L. Rubio de Francia, 'Maximal and singular integral operators via Fourier transform estimates', *Invent. Math.* **84** (1984), 541–562.

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

[24] H. Furstenberg, Problems Session, Conference on Ergodic Theory and Applications University of New Hampshire, Durham, NH, June 1982.

[25] N. Frantzikinakis, 'Some open problems on multiple ergodic averages', *Bull. Hellenic Math. Soc.* **60** (2016), 41–90.

[26] N. Frantzikinakis and B. Kra, 'Polynomial averages converge to the product of integrals', *Israel J. Math.* **148** (2005), 267–276.

[27] H. Furstenberg, 'Ergodic behavior of diagonal measures and a theorem of Szemeredi on arithmetic progressions', *J. Anal. Math.* **31** (1977), 204–256.

[28] H. Furstenberg, 'Nonconventional ergodic averages', in *The Legacy of John von Neumann* (Amer. Math. Soc., Providence, RI, 1990), 43–56.

[29] H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory* (Princeton University Press, 1981).

[30] H. Furstenberg and B. Weiss, 'A mean ergodic theorem for $\frac{1}{N} \sum_{n=1}^{N} f(T^n x) g(T^{n^2} x)$', in *Convergence in Ergodic Theory and Probability* (de Gruyter, Berlin, 1996), 193–227.

[31] B. Host and B. Kra, 'Non-conventional ergodic averages and nilmanifolds', *Ann. Math.* **161** (2005), 397–488.

[32] B. Host and B. Kra, 'Convergence of polynomial ergodic averages', *Israel J. Math.* **149** (2005), 1–19.

[33] A. D. Ionescu, A. Magyar, E. M. Stein and S. Wainger, 'Discrete Radon transforms and applications to ergodic theory', *Acta Math.* **198** (2007), 231–298.

[34] A. D. Ionescu and S. Wainger, 'L^p boundedness of discrete singular Radon transforms', *J. Amer. Math. Soc.* **19**(2) (2005), 357–383.

[35] H. Iwaniec and E. Kowalski, *Analytic Number Theory* vol. **53** (Amer. Math. Soc. Colloquium Publications, Providence RI, 2004).

[36] A. D. Ionescu, Á.s Magyar, M. Mirek and T. Z. Szarek, 'Polynomial averages and pointwise ergodic theorems on nilpotent groups', *Invent. Math.* **231** (2023), 1023–1140.

[37] R. L. Jones, J. M. Rosenblatt and M. Wierdl, 'Oscillation inequalities for rectangles', *Proc. Amer. Math. Soc.* **129**(5) (2001), 1349–1358.

[38] R. L. Jones, A. Seeger and J. Wright, 'Strong variational and jump inequalities in harmonic analysis', *Trans. Amer. Math. Soc.* **360**(12) (2008), 6711–6742.

[39] A. A. Karatsuba, *Basic Analytic Number Theory* (Springer-Verlag, Berlin, 1993). Translated from the second (1983) Russian edition and with a preface by Melvyn B. Nathanson.

[40] A. Y. Khintchin, 'Zur Birkhoff's Lb'sung des Ergodensproblems', *Math. Ann.* **107** (1933), 485–488.

[41] B. Krause, M. Mirek and T. Tao, 'Pointwise ergodic theorems for non-conventional bilinear polynomial averages', *Ann. Math.* **195**(3) (2022), 997–1109.

[42] P. LaVictoire, 'Universally L^1-bad arithmetic sequences', *J. Anal. Math.* **113**(1) (2011), 241–263.

[43] A. Leibman, 'Convergence of multiple ergodic averages along polynomials of several variables', *Israel J. Math.* **146** (2005), 303–315.

[44] A. Magyar, E. M. Stein and S. Wainger, 'Discrete analogues in harmonic analysis: spherical averages', *Ann. Math.* **155** (2002), 189–208.

[45] A. Magyar, E. M. Stein and S. Wainger, 'Maximal operators associated to discrete subgroups of nilpotent Lie groups', *J. Anal. Math.* **101**(1) (2007), 257–312.

[46] M. Mirek, '$\ell^p(\mathbb{Z}^d)$-estimates for discrete Radon transform: square function estimates', *Anal. PDE* **11**(3) (2018), 583–608.

[47] M. Mirek, W. Slomian and T. Szarek, 'Some remarks on oscillation inequalities', *Ergodic Theory and Dynam.* Systems, online version **29** (November 2022), 1–30.

[48] M. Mirek, E. M. Stein and B. Trojan, '$\ell^p(\mathbb{Z}^d)$-estimates for discrete operators of Radon types I: maximal functions and vector-valued estimates', *J. Funct. Anal.* **277**(8) (2019), 2471–2521.

[49] M. Mirek, E. M. Stein and B. Trojan, '$\ell^p(\mathbb{Z}^d)$-estimates for discrete operators of Radon type: variational estimates', *Invent. Math.* **209**(3) (2017), 665–748.

[50] M. Mirek, E. M. Stein and P. Zorin-Kranich, 'Jump inequalities via real interpolation', *Math. Ann.* **376**(1–2) (2020), 797–819.

[51] M. Mirek, E. M. Stein and P. Zorin-Kranich. 'A bootstrapping approach to jump inequalities and their applications', *Analysis & PDE* **13**(2) (2020), 527–558.

[52] M. Mirek, E. M. Stein and P. Zorin-Kranich, 'Jump inequalities for translation-invariant operators of Radon type on $\mathbb{Z}^d$', *Advances in Mathematics* **365**(107065) (2020), 57.

[53] M. Mirek, T. Z. Szarek and J. Wright, 'Oscillation inequalities in ergodic theory and anaysis; one parameter and multi-parameter perspectives', *Rev. Mat. Iberoam.* **38**(7) (2022), 2249–2284.

[54] M. Mirek and B. Trojan, 'Discrete maximal functions in higher dimensions and applications to ergodic theory', *Amer. J. Math.* **138**(6) (2016), 1495–1532.

[55] L. B. Pierce, 'On superorthogonality', *J. Geom. Anal.* **31** (2021), 7096–7183.

[56] W. Sierpiński, 'Sur la valeur asymptotique d'une certaine somm', *Bull. Intl. Acad. Polonaise des Sci. et des Lettres (Cracovie) series A* (1910), 9–11.

[57] E. M. Stein, *Harmonic Analysis* (Princeton University Press, 1993).

[58] E. M. Stein and S. Wainger' 'Discrete analogues in harmonic analysis I: ℓ^2 estimates for singular Radon transforms', *Amer. J. Math.* **121** (1999), 1291–1336.

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

[60] T. Tao, 'Equidistribution for multidimensional polynomial phases', available at Terence Tao's blog, 06 August 2015: terrytao.wordpress.com/2015/08/06/equidistribution-for-multidimensional-polynomial-phases/.

[61] T. Tao, 'Norm convergence of multiple ergodic averages for commuting transformations', *Ergodic Theory Dynam. Systems* **28** (2008), 657–688.

[62] T. Tao, 'The Ionescu–Wainger multiplier theorem and the adeles', *Mathematika* **63**(3) (2021), 557–737.

[63] I. M. Vinogradov, *The Method of Trigonometrical Sums in the Theory of Numbers* (Interscience Publishers, New York, 1954).

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

[65] M. Walsh, 'Norm convergence of nilpotent ergodic averages', *Ann. Math.* **175**(3) (2012), 1667–1688.

[66] H. Weyl, 'Über die Gibbs'sce Erscheinung und verwandte Konvergenzphenomene', *Rendiconti del Circolo Matematico di Palermo* **330** (1910), 377–407.

[67] H. Weyl, 'Über die Gleichverteilung von Zahlen mod. Eins', *Math. Ann.* **7** (1916), 313–352.

[68] T. D. Wooley, 'Nested efficient congruencing and relatives of Vinogradov's mean value theorem', *Proc. London Math. Soc. (3)* **118**(4) (2019), 942–1016.

[69] T. Wooley, 'The cubic case of the main conjecture in Vinogradov's mean value theorem', *Adv. Math.* **294** (2016), 532–561.

[70] T. D. Wooley, 'Vinogradov's mean value theorem via efficient congruencing', *Ann. Math.* **175**(3) (2012), 1575–1627.

[71] T. Ziegler, 'Universal characteristic factors and Furstenberg averages', *J. Amer. Math. Soc.* **20** (2007), 53–97.

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

[73] A. Zygmund, *Trigonometric Series* third edn. (Cambridge University Press, 2003).