POINTWISE CONVERGENCE OF POLYNOMIAL MULTIPLE ERGODIC AVERAGES ALONG THE PRIMES

RENHUI WAN

ABSTRACT. We establish pointwise almost everywhere convergence for the polynomial multiple ergodic averages

$$\frac{1}{N} \sum_{n=1}^{N} \Lambda(n) f_1(T^{P_1(n)}x) \cdots f_k(T^{P_k(n)}x)$$

as $N \to \infty$, where Λ is the von Mangoldt function, $T: X \to X$ is an invertible measure-preserving transformation of a probability space $(X, \nu), P_1, \dots, P_k$ are polynomials with integer coefficients and distinct degrees, and $f_1, \dots, f_k \in L^{\infty}(X)$. This pointwise almost everywhere convergence result can be seen as a refinement of the norm convergence result obtained in Wooley–Ziegler (Amer. J. Math, 2012) in the case of polynomials with distinct degrees.

Building on the foundational work of Krause–Mirek–Tao (Ann. of Math., 2022), Kosz–Mirek–Peluse–Wright (arXiv: 2411.09478, 2024), and Krause–Mousavi–Tao–Teräväinen (arXiv: 2409.10510, 2024), we develop a multilinear circle method for von Mangoldt-weighted (equivalently, prime-weighted) averages. This method combines harmonic analysis techniques across multiple groups with the newest inverse theorem from additive combinatorics. In particular, the principal innovations of this framework include: (i) an inverse theorem and a Weyl-type inequality for multilinear Cramér-weighted averages; (ii) a multilinear Rademacher-Menshov inequality; and (iii) an arithmetic multilinear estimate.

1. Introduction

1.1. Motivation and main result. A fundamental issue in ergodic theory is to comprehend the convergence of multilinear polynomial ergodic averages, both in norm and pointwise almost everywhere. This exploration began in the early 1930s with von Neumann's mean ergodic theorem [45] and Birkhoff's pointwise ergodic theorem [5]. In Furstenberg's work [20], multilinear ergodic averages for linear polynomials and a single transformation emerged as a natural tool for identifying recurrent points and, consequently, arithmetic progressions within subsets of integers that possess positive upper density. This approach particularly leads to the well-known Szemerédi's theorem [49] and has sparked the development of a new field now referred to as ergodic Ramsey theory. Over the past century, significant progress has been made in this area of research, which we will summarize shortly.

Let $X = (X, \nu)$ be a probability space and $T : X \to X$ is an invertible measure preserving map, meaning that $\nu(T^{-1}(E)) = \nu(E)$ for all measurable sets $E \subset X$. The triple $X = (X, \nu, T)$ is referred to as a measure-preserving system. Given complex-valued functions $f_1, \ldots, f_k \in L^{\infty}(X)$ and polynomials P_1, \ldots, P_k with integer coefficients, a scale $N \geq 1$, and a weight function $w \colon \mathbb{N} \to \mathbb{C}$, we define the weighted polynomial multiple ergodic averages by

$$A_{N,w;X}^{P_1,\dots,P_k}(f_1,\dots,f_k)(x) := \mathbb{E}_{n\in[N]}w(n)f_1(T^{P_1(n)}x)\cdots f_k(T^{P_k(n)}x), \qquad x\in X.$$
(1.1)

(See Subsection 2.1 for the definition of the notation $\mathbb{E}_{n\in[N]}$.) In this paper, we consider the problem of the pointwise almost everywhere convergence of $A_{N,\Lambda;X}^{P_1,\ldots,P_k}$, where Λ is the von

²⁰²⁰ Mathematics Subject Classification. 37A30, 37A46, 42A45, 42A85, 43A25, 11B30.

Key words and phrases. Ergodic averages, pointwise convergence, variational estimates, inverse theorem, circle method, adelic integers, Fourier method.

Mangoldt function (see, e.g., [28, Chapter 1]) given by

$$\Lambda(n) = \begin{cases} \log p, & n \text{ is a power of a prime } p, \\ 0, & \text{otherwise.} \end{cases}$$

1.1.1. Unweighted ergodic average. By setting w = 1 in (1.1), we obtain the unweighted ergodic averages

$$A_{N,1:X}^{P_1,\dots,P_k}(f_1,\dots,f_k)(x) := \mathbb{E}_{n\in[N]}f_1(T^{P_1(n)}x)\cdots f_k(T^{P_k(n)}x). \tag{1.2}$$

The norm convergence and pointwise almost everywhere convergence properties of (1.2) as $N \to \infty$ have been studied by numerous mathematicians.

The $L^2(X)$ norm convergence of the averages (1.2) and more general cases is well understood, largely due to the pioneering work of Walsh [55]. Before Walsh's contributions, there were many important works aimed at establishing $L^2(X)$ norm convergence for (1.2). This includes fundamental work for linear polynomials due to Host–Kra [25] and, independently, Ziegler [58]. We also direct the reader to Leibman's appendix in [4], which provides a comparative analysis of the characteristic factor constructions introduced by Host–Kra [25] and Ziegler [58], demonstrating their equivalence. The case of general polynomials was addressed by Leibman [33], Frantzikinakis–Kra [19], and Host–Kra [26]. In addition, the limiting function for (1.2) can be identified through the theory of Host–Kra factors [25] and equidistribution on nilmanifolds [33]. For the more general averages

$$\mathbb{E}_{n \in [N]} f_1(T_1^{P_1(n)} x) \cdots f_k(T_k^{P_k(n)} x), \qquad x \in X,$$
(1.3)

where $T_1, \ldots, T_k : X \to X$ are commuting invertible measure-preserving transformations, Chu–Frantzikinakis–Host [12] established $L^2(X)$ norm convergence for (1.3) under the assumption that the polynomials P_1, \ldots, P_k have different degrees; we refer to [50, 1, 24] for the case of linear polynomials. Walsh [55] finally demonstrated norm convergence of (1.3) in the general cases, including scenarios involving noncommutative transformations T_1, \ldots, T_k generating a nilpotent group. On the other hand, identifying the limit for general polynomial ergodic averages (1.3) remains a well-known open problem in ergodic theory. Recently, a significant advancement was made by Frantzikinakis and Kuca [18], who identified the $L^2(X)$ limit for (1.3) in the case of commuting transformations and linearly independent polynomials.

The current understanding of pointwise a.e. convergence of multilinear polynomial ergodic averages is less developed. Although pointwise a.e. convergence is the most intuitive form of convergence, it is quite subtle and can differ markedly from norm convergence, which makes investigating pointwise convergence issues particularly challenging. According to the dominated convergence theorem, pointwise a.e. convergence of the averages (1.2) implies their norm convergence for all $f_1, \ldots, f_k \in L^{\infty}(X)$ on a probability space (X, ν) . Thus, from a theoretical perspective, the pointwise a.e. convergence can be seen as a strict refinement of the norm convergence. The question of pointwise a.e. convergence is the subject of the celebrated Furstenberg-Bergelson-Leibman conjecture [3, Section 5.5]. In two notable works [7] and [6], Bourgain established the pointwise a.e. convergence of (1.2) for the case where k=1 and the case where k=2 (P_1, P_2 are linear), respectively. For the related results on the variational inequality (see Section 2 for its definition) yielding the pointwise a.e. convergence, we refer to [9, 40] and the references therein. Krause, Mirek and Tao made significant contributions in their seminal work [31], proving the pointwise a.e. convergence of (1.2) for the bilinear case with polynomials of the form (n, P(n)) where deg $P \geq 2$. Very recently, Kosz, Mirek, Peluse and Wright in their significant work [30] achieved the pointwise a.e. convergence of (1.3) for arbitrary $k \in \mathbb{Z}_+$ and polynomials P_1, \ldots, P_k with distinct degrees. This result gives an affirmative answer to a question posed by Bergelson [2, Question 9] in the case of polynomials with distinct degrees.

1.1.2. Prime-weighted ergodic averages. The objective of this paper is to prove the pointwise a.e. convergence for the weighted polynomial multiple ergodic averages $A_{N,\Lambda;X}^{P_1,\dots,P_k}$. By a routine computation, the pointwise a.e. convergence for this weighted averages are consistent with that for the prime-weighted averages given by¹

$$\frac{1}{N/\log N} \sum_{n \le N} f_1(T^{P_1(p)}x) \cdots f_k(T^{P_k(p)}x), \qquad x \in X.$$
 (1.4)

Throughout this paper we will do not distinguish these two weighted averages. The norm convergence of polynomial ergodic prime-weighted averages (1.4) is now known for any number of polynomial iterates, thanks to the work of Wooley–Ziegler [57] (see Frantzikinakis–Host–Kra [17] for the case of linear polynomials).

However, similar to the unweighted averages, the problem of pointwise a.e. convergence of ergodic averages along the primes is more involved. The case of k=1 with linear polynomials was established by Bourgain [8] and Wierdl [56] (with the latter work allowing L^q functions for any q>1), while the case of an arbitrary single polynomial iterate was handled by Nair [43, 44]. We refer to [41, 37, 54] for the related works on the variational inequality. For the bilinear case, by employing the method from Krause-Mirek-Tao [31] and the generalized von Neumann theorem established by Teräväinen [53] (which handles a broad class of weight functions), Krause-Mousavi-Tao-Teräväinen [32] established the pointwise a.e. convergence for polynomials of the form (n, P(n)) with deg $P \geq 2$. Beyond this specific bilinear result, to the best of our knowledge, no other pointwise a.e. convergence results exist for the multilinear averages $A_{N,\Lambda;X}^{P_1,\dots,P_k}$ or (1.4). We also mention that the problem of pointwise a.e. convergence of weighted ergodic averages with more than one iterate was discussed by Frantzikinakis [16].

- 1.1.3. Statement of our main result. Motivated by the studies on
 - the pointwise a.e. convergence of multilinear unweighted ergodic averages [31, 30],
 - the pointwise a.e. convergence of bilinear prime-weighted ergodic averages [32],
 - and the norm convergence of multiple prime-weighted ergodic averages [57],

we address the following question:

Question 1. Let $k \in \mathbb{Z}_+$ and (X, ν, T) be a measure-preserving system.² Let P_1, \ldots, P_k be polynomials with integer coefficients and distinct degrees. Is it true that for any functions $f_1, \ldots, f_k \in L^{\infty}(X)$ the multilinear polynomial ergodic averages $A_{N,\Lambda;X}^{P_1,\ldots,P_k}$ converge pointwise almost everywhere on X as $N \to \infty$?

This question extends [2, Question 9] (in the single transformation case) posed by Bergelson to the prime-weighted setting, while generalizing a question from Frantzikinakis's survey [16, Problem 12] to polynomial cases. We now state the main result of this paper.

Theorem 1.1 (Main result). Let the notation and hypotheses be as stated in Question 1. Then, for any $f_1, \ldots, f_k \in L^{\infty}(X)$, the following results hold.

- (i) (Pointwise ergodic theorem) The averages $A_{N,\Lambda;X}^{P_1,\dots,P_k}(f_1,\dots,f_k)$ converge pointwise a.e. as $N\to\infty$.
- (ii) (Variational ergodic theorem) If r > 2 and $0 < q < \infty$, then one has $\left\| \left(A_{N,\Lambda;X}^{P_1,\dots,P_k}(f_1,\dots,f_k) \right)_{N\in\mathbb{D}} \right\|_{L^q(X;\mathbf{V}^r)} \lesssim_{P_1,\dots,P_k,\lambda,r,q} \|f_1\|_{L^\infty(X)} \cdots \|f_k\|_{L^\infty(X)},$ where $\mathbb{D} = \{\lambda_n \in \mathbb{N} : n \in \mathbb{N}\} \subset [1,+\infty)$ is λ -lacunary, i.e., $\inf_{n\in\mathbb{N}} \frac{\lambda_{n+1}}{\lambda_n} \geq \lambda$.

 $^{^{1}}$ All sums and products over the symbol p will be understood to be over primes; other sums will be understood to be over positive integers unless otherwise specified.

²Throughout this paper, all measure-preserving systems are assumed to have finite measure, while all σ -finite measure-preserving systems will be explicitly stated as such.

Comments on Theorems 1.1 are given as follows:

- (1) The conclusion from part (i) provides an affirmative answer to *Question* 1 and establishes the pointwise a.e. convergence of the prime-weighted averages (1.4). Additionally, with necessary adjustments, our method remains applicable to the unweighted averages (1.2) in the case of polynomials with distinct degrees.
- (2) The primary significance of the theorem is in the multilinear case $k \geq 2$, whereas the variational inequality for the linear case k = 1 (including scenarios involving σ -finite measure-preserving system and the full variational inequality) has already been well thoroughly resolved. Accordingly, the following auxiliary theorems consider only $k \geq 2$.
- (3) By combining our results and the dominated convergence theorem, we can immediately deduce the norm convergence of the averages $A_{N,\Lambda;X}^{P_1,\dots,P_k}(f_1,\dots,f_k)$, which was originally proven by Wooley–Ziegler [57]. Furthermore, we remove the restrictions (k=2) and special polynomials of the form (n,P(n)) with deg $P\geq 2$ required in Krause–Mousavi–Tao–Teräväinen [32], and obtain the pointwise a.e. convergence for arbitrary $k\in\mathbb{Z}_+$ and polynomials of distinct degrees for the case of the measure-preserving system. This, however, comes at the cost of imposing $f_1,\dots,f_k\in L^\infty(X)$. (In fact, we can also obtain part (i) for the general σ -finite measure-preserving system with the functions f_1,\dots,f_k in some bigger Lebesgue spaces. See Theorem 1.2 below for the details.)
- (4) The conclusion from part (ii) in Theorem 1.1 serves as the central result of the above theorem and straightforwardly leads to the result from part (i). The requirement that r > 2 is necessary, as variational estimates cannot be established for $r \le 2$ except for certain specific operators (see, for example, [13]).
- 1.2. Reduction to the integer shift system. In the study of pointwise convergence problems, the integer shift system emerges as the most significant dynamical system, as exemplified below.

Example 1 (integer shift system). The integer shift system $(\mathbb{Z}, \nu_{\mathbb{Z}}, T_{\mathbb{Z}})$ is the set of integers \mathbb{Z} equipped with counting measure $\nu_{\mathbb{Z}}$ and the shift $T_{\mathbb{Z}}(x) := x - 1$. The averages $A_{N,\Lambda;X}^{P_1,\dots,P_k}$ with $T = T_{\mathbb{Z}}$ and $X = \mathbb{Z}$ can be expressed as

$$A_{N,\Lambda;\mathbb{Z}}^{P_1,\dots,P_k}(f_1,\dots,f_k)(x) := \mathbb{E}_{n\in[N]}\Lambda(n)f_1(x-P_1(n))\cdots f_k(x-P_k(n)), \qquad x\in\mathbb{Z}.$$
 (1.6)

The associated truncated version can be given by

$$\tilde{A}_{N,\Lambda;\mathbb{Z}}^{P_1,\dots,P_k}(f_1,\dots,f_k)(x) := \frac{1}{\lfloor N \rfloor} \sum_{n \in J_N} \Lambda(n) f_1(x - P_1(n)) \cdots f_k(x - P_k(n)), \quad x \in \mathbb{Z}, \quad (1.7)$$

where $J_N := [N] \setminus [N/2]$. We will often abbreviate

$$A_{N,\Lambda}^{\mathcal{P}}:=A_{N,\Lambda;\mathbb{Z}}^{P_1,\dots,P_k}\qquad\text{and}\qquad \tilde{A}_{N,\Lambda}^{\mathcal{P}}:=\tilde{A}_{N,\Lambda;\mathbb{Z}}^{P_1,\dots,P_k}.$$

To prove part (ii) in Theorem 1.1, we shall use the Calderón transference principle to reduce the general measure-preserving system to the shift system $(\mathbb{Z}, \nu_{\mathbb{Z}}, T_{\mathbb{Z}})$. Before presenting the theorem for the shift system (a special σ -finite measure-preserving system), we first establish a general result applicable to all σ -finite measure-preserving systems. However, a key departure from the previous works [31, 30] on unweighted averages is our requirement that all functions f_1, \ldots, f_k simultaneously occupy two distinct Lebesgue spaces in the multilinear case $(k \geq 2)$. To prove Theorem 1.1, it suffices to show the following theorem.

³From an application perspective, particularly in combinatorics and related fields where the statistical properties of ergodic averages for totally ergodic systems are significant, the general framework of σ -finite measure spaces is not interesting and only finite measure spaces hold importance. See comment 14 about [30, Theorem 1.12] for the details.

Theorem 1.2 (Auxiliary theorem I). Let $k \geq 2$, (X, ν, T) be a σ -finite measure-preserving system, and let $\mathcal{P} = \{P_1, \dots, P_k\}$ be a family of polynomials with integer coefficients and distinct degrees. For any $2 \leq q < \infty$, there exists a set $\{q_{jv}\}_{(j,v) \in [k] \times \{1,2\}} \subset (1,\infty)$ with $q_{j1} \neq q_{j2}$ and $\frac{1}{q_{j1}} + \sum_{i \in [k] \setminus \{j\}} \frac{1}{q_{i2}} = \frac{1}{q}$ for all $j \in [k]$ such that for any r > 2, we have

$$\| \left(A_{N,\Lambda;X}^{P_1,\dots,P_k}(f_1,\dots,f_k) \right)_{N \in \mathbb{D}} \|_{L^q(X;\mathbf{V}^r)}$$

$$\lesssim_{\mathcal{P},\lambda,r,q,q_{11},q_{12},\dots,q_{k1},q_{k2}} \sum_{j \in [k]} \left(\| f_j \|_{L^{q_{j1}}(X)} \prod_{i \in [k] \setminus \{j\}} \| f_i \|_{L^{q_{i2}}(X)} \right)$$

$$(1.8)$$

for any $f_j \in \bigcap_{v=1,2} L^{q_{jv}}(X)$ with $j \in [k]$, where $\mathbb{D} \subset [1,\infty)$ is λ -lacunary.

The inequality (1.5) follows directly from (1.8) because $L^{\infty}(X) \subset (\bigcap_{0 \le s \le \infty} L^s(X))$ holds for any finite measure space X. Moreover, the parameters $\{q_{jv}\}_{(j,v)\in[k]\times\{1,2\}}$ are not uniquely determined; in fact, we prove a stronger result, as detailed in comment (1) below Theorem 1.3.

In view of the Calderón transference principle [11] (or, more precisely, the arguments from [31, Proposition 3.2 (ii)] or [29, Theorem 1.6]), it suffices to work with the integer shift system. In fact, the Calderón transference principle allows us to transfer the quantitative estimates (1.8) from the integer shift system to corresponding estimates for $A_{N,\Lambda;X}^{\mathcal{P}}$ in the abstract measure-preserving system (X, ν, T) . This will allow us to employ Fourier methods on \mathbb{Z} and utilize the algebraic structure of \mathbb{Z} , which are generally not available in abstract measure-preserving systems. It is important to note that the Calderón transference principle [11] only transfers quantitative bounds that imply pointwise a.e. convergence, but does not transfer pointwise a.e. convergence itself. Therefore, we will focus on proving quantitative bounds for $A_{N,\Lambda}^{\mathcal{P}}$ or $\tilde{A}_{N,\Lambda}^{\mathcal{P}}$ in the integer shift system, rather than pointwise convergence on \mathbb{Z} . Due to specific technical considerations, we will restrict our attention to the truncated averages $\tilde{A}_{N,\Lambda}^{\mathcal{P}}$.

After these reductions, Theorem 1.2 now follows from the theorem below.

Theorem 1.3 (Auxiliary theorem II). Let $k \geq 2$, and let $\mathcal{P} = \{P_1, \dots, P_k\}$ be a family of polynomials with integer coefficients and distinct degrees. For any $2 \leq q < \infty$, there exists a set $\{q_{jv}\}_{(j,v)\in[k]\times\{1,2\}}\subset (1,\infty)$ with $q_{j1}\neq q_{j2}$ and $\frac{1}{q_{j1}}+\sum_{i\in[k]\setminus\{j\}}\frac{1}{q_{i2}}=\frac{1}{q}$ for all $j\in[k]$ such that for any r>2 and for any $f_j\in\cap_{v=1,2}\ell^{q_{jv}}(\mathbb{Z})$ with $j\in[k]$, we have

$$\| \left(\tilde{A}_{N,\Lambda}^{\mathcal{P}}(f_1, \dots, f_k) \right)_{N \in \mathbb{D}} \|_{\ell^q(\mathbb{Z}; \mathbf{V}^r)}$$

$$\lesssim_{\mathcal{P}, \lambda, r, q, q_{11}, q_{12}, \dots, q_{k1}, q_{k2}} \sum_{j \in [k]} \left(\| f_j \|_{\ell^{q_{j1}}(\mathbb{Z})} \prod_{i \in [k] \setminus \{j\}} \| f_i \|_{\ell^{q_{i2}}(\mathbb{Z})} \right),$$

$$(1.9)$$

where $\tilde{A}_{N,\Lambda}^{\mathcal{P}}$ is given by (1.7), and $\mathbb{D} \subset [1,+\infty)$ is λ -lacunary.

Following the arguments yielding [31, Proposition 3.2 (ii) and (iii)], we can obtain Theorem 1.2 from Theorem 1.3. Thus, it remains to prove Theorem 1.3, which is the objective of much of the remainder of the paper. Let us make some comments on Theorems 1.3.

- (1) We actually establish the following strengthened version of Theorem 1.3 (see Subsection 4.1 and the hypotheses in Theorem 4.1). Let $\{I_n(d)\}_{n\in\mathbb{N}}$ (defined by (4.1) below) be the partition of the interval $[2,\infty)$, where $d:=\max\{\deg P_i:i\in[k]\}$. For $n\in\mathbb{N}$ and $q\in I_n(d)$, let $\{q_{jv}\}_{(j,v)\in[k]\times\{1,2\}}\subset (1,\infty)$ satisfying $\frac{1}{q_{j1}}+\sum_{i\in[k]\setminus\{j\}}\frac{1}{q_{i2}}=\frac{1}{q}$ with $j\in[k]$ and $q_{11},q_{21},\ldots,q_{k1}\in I_n(d)$ (noting that $q_{j1}\neq q_{j2}$ for all $j\in[k]$ based on these assumptions). Then, inequality (1.9) holds for any $f_j\in\cap_{v=1,2}\ell^{q_{jv}}(\mathbb{Z})$ with $j\in[k]$.
- (2) The restrictions on the parameters $q, \{q_{jv}\}_{(j,v)\in[k]\times\{1,2\}}$ and the functions f_1, \ldots, f_k are essential in proving the major arcs estimates for the low-frequency case at a large scale (see Section 6). More precisely, these restrictions are due to the application of the norm interchanging trick (see (6.13)) and the *p*-adic estimates (see (6.26) and (6.27)).

- Furthermore, we refer to *Remark* 6 below Proposition 4.6 for the justification of the main result in [31], derived using our results.
- (3) From the establishment of the optimal inequality, it is a natural and interesting problem to consider whether the aforementioned restrictions can be relaxed to match the conditions required for the multilinear unweighted averages. However, we do not pursue this direction here, as Theorem 1.3 is sufficient for us to establish Theorem 1.1.
- 1.3. Overview of the proof. Since the averages considered in this paper are multilinear, the classical circle method for the linear case cannot be applied, as discussed in [31, 30]. In fact, our proof will primarily relies on the framework established in [31], while also drawing inspiration from [30] and [32]. However, when we consider the pointwise convergence problem of multilinear averages with the von Mangoldt weight, we encounter new difficulties.
- 1.3.1. Multilinear circle method for unweighted averages. We start by discussing some limitations of two crucial multilinear circle methods presented in [31] and [30], which deal with the unweighted averages $A_{N,1;X}^{P_1,\dots,P_k}$. For specific details about their methods, we refer to [31, Subsection 1.4] and [30, Subsection 1.6].
 - The method from [31] is confined to the bilinear case, the single transformation case, and the polynomials of the form (n, P(n)) where $\deg P \geq 2$. These limitations are critical in various parts of their proof, such as deriving the alternate inverse theorems from Peluse's inverse theorem, establishing the bilinear Rademacher-Menshov inequality, transitioning the variational inequality from integers to adelic integers, and proving a desired bilinear arithmetic estimate (see [31, Theorem 9.9]) based on the trivial (yet significant) inequality

$$||A_{\mathbb{Z}_p}^{n,P(n)}||_{L^1(\mathbb{Z}_p)\times L^{\infty}(\mathbb{Z}_p)\to L^{\infty}(\mathbb{Z}_p)} \le 1.$$

$$(1.10)$$

See Subsection 2.3 below (or [31, Section 4]) for the definition of the p-adic integers \mathbb{Z}_p . It is noteworthy that the linear polynomial in (1.10) is indispensable. Although the method from [31] has its limitations, its main advantage is that the decay $2^{-cl} + \langle \text{Log } N \rangle^{-C_1}$ (where c is small and C_1 is sufficiently large) for the minor arcs estimates (see [31, Theorem 5.12]) does not require the same level of rigor as that required in [30].

- The method outlined in [30] appears to surpass that of [31], as it eliminates all previously mentioned restrictions and requires only polynomials with distinct degrees. However, it relies on the minor arcs estimates (see [30, Theorem 6.1]) with a polynomial decay $2^{-cl} + N^{-c}$ (with small c) to effectively utilize the metric entropy argument (see [30, Subsection 7.3.6 and Remark 7.94]). Unfortunately, the weighted averaging operator considered in the present paper does not exhibit such a favorable decay (precisely, $2^{-cl} + \exp(-c \log^{1/C_0} N)$) with small c and large C_0 , see Theorem 3.1), even with the application of the new Ionescu-Wainger multiplier theorem (see Theorem 2.1), rendering the approach in [30] directly inapplicable. This is the main reason why we are not able to achieve the case of arbitrary commuting measure-preserving transformations as in [30].
- 1.3.2. Multilinear circle method for averages with von Mangoldt weight. To our knowledge, the general multilinear circle method, which addresses the weighted averages $A_{N,\Lambda;X}^{P_1,\dots,P_k}$, remains unexplored, aside from the bilinear case in [32]. In particular, the limitations of the method in [31] also apply to the method in [32], as the latter is heavily reliant on the former. Based on the discussion in Subsection 1.3.1, we cannot directly utilize the two methods mentioned. Instead, we will develop a multilinear circle method for the weighted averages $A_{N,\Lambda;X}^{P_1,\dots,P_k}$, building on the

approaches in [31, 30, 32] while introducing some new novelties. We first present these new novelties.

- We establish an inverse theorem and a version of Weyl's inequality (termed the multilinear Weyl inequality) for multilinear Cramér-weighted averages (see Section 3) involving general polynomials with distinct degrees (see Theorems 3.1 and 3.3). The latter provides the desired minor arcs estimates. Additionally, in the specific bilinear case involving the polynomial pair (P_1, P_2) , where P_1 is linear and P_2 has a degree of at least 2, the decay achieved in this multilinear Weyl inequality improves upon the result in [32, Theorem 3.2].
- We develop a multilinear Rademacher-Menshov inequality (see Lemma 5.1) alongside an arithmetic multilinear estimate (see Lemma 6.1). The former is employed to prove the major arcs estimates for the low-frequency case at a small scale, while the latter addresses the major arcs estimates for the low-frequency case at a large scale. Moreover, these results can be viewed as generalizations of the bilinear Rademacher-Menshov inequality and the arithmetic bilinear estimate established in [31, Lemma 8.1 and Theorem 9.9], but they cannot be trivially derived from the arguments presented in that work.

We now outline the proof of Theorem 1.3. Utilizing the little Gowers norm estimate (refer to (4.10) below) and the generalized von Neumann theorem (see Lemma 2.3), we reduce the proof of Theorem 1.3 to demonstrating Theorem 3.1 and Theorem 4.1, which provide the minor arcs estimates and major arcs estimates for the Cramér-weighted averages, respectively.

• Sketch of the proof of Theorem 3.1: We prove the Cramér-weighted inverse theorem (see Theorem 3.3) and subsequently prove Theorem 3.1 by combining Theorem 3.3, the weighted L^p improving estimates (see Lemma 3.8), and the arguments that address unweighted averages, such as the Hahn-Banach theorem (see [31, Lemma 6.9]), dual arguments, and the new Ionescu-Wainger multiplier theorem 4 (see Theorem 2.1).

To obtain this Cramér-weighted inverse theorem (see Subsection 3.1.2), we combine a reduction argument with a new unweighted inverse theorem whose proof needs several crucial techniques, including the alternate inverse theorem in [31], the Peluse's inverse theorem [47, Theorem 3.3], and the little Gowers norm estimates related to the Cramér and Heath-Brown approximants (see Subsection 2.6). Notably, the lower bound of δ (see (3.22)) in the Cramér-weighted inverse theorem will determine the decay in the multilinear Weyl inequality (3.4), but the polynomial decay required by the approach in [30] has not been achieved.

• Sketch of the proof of Theorem 4.1: We employ both arithmetic and continuous dyadic decompositions to reduce the proof of Theorem 4.1 to demonstrating Propositions 4.4, 4.5, and 4.6, which correspond to the high-frequency case, the low-frequency case at a small scale, and the low-frequency case at a large scale, respectively.

Proposition 4.4 will be established by proving a multilinear Weyl inequality in the continuous setting (also referred to as the Sobolev smoothing inequality) given by Theorem A.1. This is achieved by combining a transference trick with [30, Theorem 6.1]. For the proof of the low-frequency case, however, the metric entropy argument presented in [30] is not applicable, as the multilinear Weyl inequality in this context does not exhibit the polynomial decay. Instead, we develop a nontrivial variant of the method from [31].

⁴While the original Ionescu-Wainger multiplier theorem (see [27, 38, 51]) could alternatively be employed, we opt for the new version in this work to achieve notational simplicity and to enable direct comparison with the corresponding arguments presented in [30].

The proof of Proposition 4.5 relies on establishing a multilinear Rademacher–Menshov inequality (see Lemma 5.1), which we derive by combining an inductive argument with the bilinear version obtained in [31].

For the proof of Proposition 4.6, we utilize the quantitative Shannon sampling theorem (see Lemma B.1) and the norm interchanging trick (see (6.13)) to reduce the matter to demonstrating the arithmetic multilinear estimate (see Lemma 6.1) and the p-adic estimates (see (6.25)). In fact, the restrictions on the functions f_1, \ldots, f_k and the parameters $q, \{q_{jv}\}_{(j,v)\in[k]\times\{1,2\}}$ in Theorem 1.3 originate entirely from the techniques used to establish these estimates. Moreover, unlike the arithmetic bilinear estimate in [31], our arithmetic multilinear estimate requires stronger constraints on the parameters q, r. This is due to the multilinear case, where no additional structural conditions are imposed on the polynomials (beyond their differing degrees). For details, see the hypotheses in Theorem 4.1 and Lemma 6.1.

1.4. Organization. In Section 2, we summarize our main notation and compile several key theorems and lemmas. Section 3 presents a multilinear Weyl inequality for Cramér-weighted averages which yields the desired minor arcs estimates. In Section 4, we reduce the proof of Theorem 1.3 to showing Propositions 4.4, 4.5 and 4.6. These propositions correspond to the major arcs estimates for the high-frequency case, the low-frequency case at a small scale, and the low-frequency case at a large scale respectively. We then proceed to prove Proposition 4.4. In Sections 5 and 6, we establish Propositions 4.5 and 4.6. Appendix A provides a multilinear Weyl inequality in the continuous setting, which is utilized in the proof of Proposition 4.4. Appendix B includes the definition of the sampling map \mathcal{S} and the essential quantitative Shannon sampling theorem, both of which are instrumental in proving Proposition 4.6.

2. NOTATION AND PRELIMINARIES

2.1. **Basic notation.** Throughout the paper C>0 is an absolute constant that may change from occurrence to occurrence. For any two quantities A and B, we will write $A \lesssim B$ to denote $A \leq CB$ for some absolute constant C. The notation A = B + O(X) means $|A - B| \lesssim X$. If we need the implied constant C to depend on additional parameters, we will denote this by subscripts. If both $A \lesssim B$ and $B \lesssim A$ hold, we use $A \sim B$. To abbreviate the notation we will sometimes permit the implied constant to depend on certain fixed parameters when the issue of uniformity with respect to such parameters is not of relevance.

We denote the prime numbers by $\mathbb{P} := \{2, 3, 5, \ldots\}$, the positive integers by $\mathbb{Z}_+ := \{1, 2, \ldots\}$ and the natural numbers by $\mathbb{N} := \mathbb{Z}_+ \cup \{0\}$. For any $x \in \mathbb{R}$, $\lfloor x \rfloor$ denotes the largest integer smaller than x. For any N > 0, we use [N] to denote the discrete interval $\{n \in \mathbb{Z}_+ : n \leq N\}$, let $\mathbb{N}_{\leq N} := [N] \cup \{0\}$ and $[\pm N] := [-N, N] \cap \mathbb{Z}$. For $q \in \mathbb{Z}_+$ and $a = (a_1, \ldots, a_k) \in \mathbb{Z}^k$ with $k \in \mathbb{Z}_+$, we denote by (a, q) the greatest common divisor of a and q; that is, the largest $d \in \mathbb{Z}_+$ that divides q and all the components a_1, \ldots, a_k . Clearly, any vector in \mathbb{Q}^k (the direct product of k copies of the rational numbers \mathbb{Q}) has a unique representation as a/q with $q \in \mathbb{Z}_+$, $a \in \mathbb{Z}^k$ and (a, q) = 1. We let $[q]^\times := \{b \in [q] : (b, q) = 1\}$ denote the elements of [q] that are coprime to q. Additionally, if \mathbb{R} is a commutative ring, we use \mathbb{R}^\times to denote the multiplicatively invertible elements of \mathbb{R} .

We use the averaging notation $\mathbb{E}_{n\in S}f(n):=\frac{1}{\#S}\sum_{n\in S}f(n)$ for any finite non-empty set S, where #S denotes the cardinality of S. We also need the Euler totient function by $\varphi(q):=\#([q]^{\times})$ for any $q\in\mathbb{Z}_+$. Moreover, $\mathbb{1}_E$ denotes the indicator function of a set E, that is, $\mathbb{1}_E(x):=\mathbb{1}_{x\in E}$. We use the Japanese bracket notation $\langle x\rangle:=(1+|x|^2)^{1/2}$ for any real or complex x. All logarithms in this paper are taken base 2, and for any $N\geq 1$, we define the

 $logarithmic\ scale\ Log\ N\ of\ N\ by\ the\ formula$

$$Log N := |log N|,$$

thus Log N is the unique natural number such that $2^{\text{Log }N} \leq N < 2^{\text{Log }N+1}$. Throughout this paper we fix a cutoff function $\eta \colon \mathbb{R} \to [0,1]$ that is a smooth even function supported on [-1,1] that equals one on [-1/2,1/2]. For any $k \in \mathbb{Z}$, we define $\eta_{\leq k} \colon \mathbb{R} \to [0,1]$ by

$$\eta_{\leq \mathbf{k}}(\xi) := \eta(\xi/2^{\mathbf{k}}). \tag{2.1}$$

2.2. Functional spaces. All vector spaces considered in this paper are defined over the complex field \mathbb{C} . Given a measure space (X, ν) , we define $L^0(X)$ as the space of all ν -measurable complex-valued functions on X, identifying functions that are equal almost everywhere with respect to ν . For $q \in (0, \infty)$, the subspace of functions in $L^0(X)$ whose absolute value raised to the q-th power is integrable is denoted by $L^q(X)$; the space $L^\infty(X)$ consists of all essentially bounded functions in $L^0(X)$. For any exponent $q \in [1, \infty]$, its Hölder conjugate $q' \in [1, \infty]$ is defined by the relation 1/q + 1/q' = 1. When the measure on X is the counting measure, we abbreviate $L^q(X)$ as $\ell^q(X)$, or simply ℓ^q .

We can extend these notation to functions that take values in a finite-dimensional normed vector space $V = (V, \|\cdot\|_V)$. For example, $L^0(X; V)$ refers to the space of measurable functions from X to V (up to almost everywhere equivalence), as well as

$$L^{p}(X;V) := \left\{ F \in L^{0}(X;V) : \|F\|_{L^{q}(X;V)} := \|\|F\|_{V}\|_{L^{q}(X)} < \infty \right\}. \tag{2.2}$$

These notions can also be extended to functions taking values in an infinite-dimensional normed vector space V, at least when V is separable. However, in most cases we will be able to work within finite-dimensional settings, or reduce to them via a standard approximation argument.

For any finite dimensional normed vector space $(B, \|\cdot\|_B)$, any sequence $(\mathfrak{a}_t)_{t\in\mathbb{I}}$ of elements of B indexed by a totally ordered set \mathbb{I} , and any exponent $1 \leq r < \infty$, the r-variation seminorm is defined by the following formula:

$$\|(\mathfrak{a}_t)_{t\in\mathbb{I}}\|_{V^r(\mathbb{I};B)} := \sup_{J\in\mathbb{Z}_+} \sup_{t_0<\dots< t_J} \Big(\sum_{j=0}^{J-1} \|\mathfrak{a}(t_{j+1}) - \mathfrak{a}(t_j)\|_B^r\Big)^{1/r}, \tag{2.3}$$

where the supremum is taken over all finite increasing sequences $\{t_j\}_{j\in[J]}\subset\mathbb{I}$, and it is set to zero by convention if \mathbb{I} is empty. As r approaches infinity, we adopt $\|(\mathfrak{a}_t)_{t\in\mathbb{I}}\|_{V^{\infty}(\mathbb{I};B)}:=\sup_{t\leq t'\in\mathbb{I}}\|\mathfrak{a}(t')-\mathfrak{a}(t)\|_{B}$.

The r-variation norm for $1 \le r \le \infty$ is defined as:

$$\|(\mathfrak{a}_t)_{t\in\mathbb{I}}\|_{\mathbf{V}^r(\mathbb{I};B)} := \sup_{t\in\mathbb{I}} \|\mathfrak{a}_t\|_B + \|(\mathfrak{a}_t)_{t\in\mathbb{I}}\|_{V^r(\mathbb{I};B)}. \tag{2.4}$$

This clearly defines a norm on the space of functions mapping from \mathbb{I} to B. When $B = \mathbb{C}$, we will use the abbreviations $V^r(\mathbb{I}; X)$ as $V^r(\mathbb{I})$ or simply V^r , and $V^r(\mathbb{I}; X)$ as $V^r(\mathbb{I})$ or V^r . If (X, μ) is a measure space, by (2.4) and (2.2), one can explicitly write

$$L^{q}(X; \mathbf{V}^{r}) = \{ F \in L^{0}(X; \mathbf{V}^{r}) : ||F||_{L^{q}(X; \mathbf{V}^{r})} := |||F||_{\mathbf{V}^{r}}||_{L^{q}(X)} < \infty \}.$$

Note that the \mathbf{V}^r norm is non-decreasing in r, and comparable to the ℓ^{∞} norm when $r = \infty$. We also observe the simple triangle inequality

$$\|(\mathfrak{a}_t)_{t\in\mathbb{I}}\|_{\mathbf{V}^r(\mathbb{I};X)} \lesssim \|(\mathfrak{a}_t)_{t\in\mathbb{I}_1}\|_{\mathbf{V}^r(\mathbb{I}_1;X)} + \|(\mathfrak{a}_t)_{t\in\mathbb{I}_2}\|_{\mathbf{V}^r(\mathbb{I}_2;X)}$$

$$(2.5)$$

whenever $\mathbb{I} = \mathbb{I}_1 \oplus \mathbb{I}_2$ is an ordered partition of \mathbb{I} . In a similar spirit, we have the bound

$$\|(\mathfrak{a}_t)_{t\in\mathbb{I}}\|_{\mathbf{V}^r(\mathbb{T}:X)} \lesssim \|(\mathfrak{a}_t)_{t\in\mathbb{I}}\|_{\ell^r(\mathbb{T}:X)}. \tag{2.6}$$

- 2.3. **Abstract harmonic analysis.** Throughout this paper, we conduct Fourier analysis across multiple groups, with special emphasis on the intricate connection between major arcs Fourier analysis on the integer group $\mathbb Z$ and low frequency Fourier analysis on the adelic integers $\mathbb A_{\mathbb Z}$. To approach this analysis systematically, we establish a framework of abstract harmonic analysis notation. We define the unit circle as $\mathbb T := \mathbb R/\mathbb Z$ and introduce the standard character $e: \mathbb T \to \mathbb C$ given by $e(\theta) := e^{-2\pi i \theta}$, where $i^2 = -1$.
- 2.3.1. Fourier transform. Let $(\mathcal{G}, +)$ be a locally compact abelian group (LCA) group equipped with a Haar measure $\nu_{\mathcal{G}}$. It is well known (see for instance [15, 48]) that every LCA group \mathcal{G} has a Pontryagin dual $\mathcal{G}^* = (\mathcal{G}^*, +)$, an LCA group with a Haar measure $\nu_{\mathcal{G}^*}$ and a pairing, i.e., a continuous bihomomorphism $\mathcal{G} \times \mathcal{G}^* \ni (x, \xi) \mapsto x \cdot \xi \in \mathbb{T}$, such that the Fourier transform $\mathcal{F}_{\mathcal{G}}: L^1(\mathcal{G}) \to C(\mathcal{G}^*)$ given by

$$\mathcal{F}_{\mathcal{G}}f(\xi) := \int_{\mathcal{G}} f(x)e(x \cdot \xi)d\nu_{\mathcal{G}}(x), \qquad \xi \in \mathcal{G}^*,$$

extends to a unitary map from $L^2(\mathcal{G})$ to $L^2(\mathcal{G}^*)$; in particular we have Plancherel's identity $\|\mathcal{F}_{\mathcal{G}}f\|_{L^2(\mathcal{G}^*)} = \|f\|_{L^2(\mathcal{G})}$ for $f \in L^2(\mathcal{G})$. Moreover, the inverse Fourier transform $\mathcal{F}_{\mathcal{G}}^{-1}$: $L^2(\mathcal{G}^*) \to L^2(\mathcal{G})$ is given by the formula

$$\mathcal{F}_{\mathcal{G}}^{-1}f(x) = \int_{\mathcal{G}^*} f(\xi)e(-x \cdot \xi)d\nu_{\mathcal{G}^*}(\xi), \quad \text{where} \quad f \in L^1(\mathcal{G}^*) \cap L^2(\mathcal{G}^*).$$

We will utilize the following concrete pairs $(\mathcal{G}, \mathcal{G}^*)$ of Pontryagin dual LCA groups:

- $(\mathcal{G}, \mathcal{G}^*) = (\mathbb{R}, \mathbb{R}).$
- $(\mathcal{G}, \mathcal{G}^*) = (\mathbb{Z}, \mathbb{T}).$
- $(\mathcal{G}, \mathcal{G}^*) = (\mathbb{Z}/Q\mathbb{Z}, \frac{1}{Q}\mathbb{Z}/\mathbb{Z})$ for some $Q \in \mathbb{Z}_+$.
- $(\mathcal{G}, \mathcal{G}^*) = (\mathbb{Z}_p := \varprojlim_j \mathbb{Z}/p^j \mathbb{Z}, \ \mathbb{Z}_p^* := \mathbb{Z}[\frac{1}{p}]/\mathbb{Z})$ for some prime $p \in \mathbb{P}$.
- $(\mathcal{G}, \mathcal{G}^*) = (\hat{\mathbb{Z}} := \prod_{p \in \mathbb{P}} \mathbb{Z}_p, \mathbb{Q}/\mathbb{Z}).$
- $(\mathcal{G}, \mathcal{G}^*) = (\mathbb{A}_{\mathbb{Z}} := \mathbb{R} \times \hat{\mathbb{Z}}, \ \mathbb{A}_{\mathbb{Z}}^* := \mathbb{R} \times \mathbb{Q}/\mathbb{Z})$. More generally, if $\mathcal{G}_1, \dots, \mathcal{G}_k$ are LCA groups with Pontryagin duals $\mathcal{G}_1^*, \dots, \mathcal{G}_k^*$, then the product $\mathcal{G} = \mathcal{G}_1 \times \dots \times \mathcal{G}_k$ with product Haar measure $\nu_{\mathcal{G}_1} \times \dots \times \nu_{\mathcal{G}_k}$ is an LCA group with Pontryagin dual $\mathcal{G}^* = \mathcal{G}_1^* \times \dots \times \mathcal{G}_k^*$ with product Haar measure $\nu_{\mathcal{G}^*} = \nu_{\mathcal{G}_1^*} \times \dots \times \nu_{\mathcal{G}_k^*}$.

For further details, we refer to [31, Section 4]. We consider the space $S(\mathcal{G}) \subset (L^1(\mathcal{G}) \cap L^{\infty}(\mathcal{G}))$ of Schwartz-Bruhat functions $f \colon G \to \mathbb{C}$, which generalizes the classical Schwartz functions on \mathbb{R} . These functions are dense in $L^q(\mathcal{G})$ for every $1 \leq q < \infty$ and are well-suited for Fourier analysis. A detailed definition of the space $S(\mathcal{G})$ can be found in [31, Section 4]. For the general case, see also [10, 46]. For $k \in \mathbb{Z}_+$, we denote by $(S(\mathcal{G}))^k$ the direct product of k copies of space $S(\mathcal{G})$.

2.3.2. Fourier multiplier operators. We now define Fourier multiplier operators. A continuous function $\mathfrak{m}: \mathcal{G}^* \to \mathbb{C}$ is said to be *smooth tempered* if $\mathfrak{m}F \in \mathcal{S}(\mathcal{G}^*)$ whenever $F \in \mathcal{S}(\mathcal{G}^*)$.

If $\mathfrak{m}: \mathcal{G}^* \to \mathbb{C}$ is a smooth tempered function, we define the linear Fourier multiplier operator $T_{\mathcal{G}}[\mathfrak{m}]: S(\mathcal{G}) \to S(\mathcal{G})$ by

$$T_{\mathcal{G}}[\mathfrak{m}]f(x) := \int_{\mathcal{G}^*} e(-x \cdot \xi)\mathfrak{m}(\xi)\mathcal{F}_{\mathcal{G}}f(\xi)d\nu_{\mathcal{G}^*}(\xi), \qquad x \in \mathcal{G}.$$

$$(2.7)$$

We refer to \mathfrak{m} as the *symbol* of $T_{\mathcal{G}}[\mathfrak{m}]$. In addition, for any finite subset $\mathfrak{Q} \subseteq \mathcal{G}^*$ we define

$$T_{\mathcal{G}}^{\mathfrak{Q}}[\mathfrak{m}]f(x) := T_{\mathcal{G}}\left[\sum_{\theta \in \mathfrak{Q}} \tau_{\theta} \mathfrak{m}\right] f(x), \qquad x \in \mathcal{G},$$
 (2.8)

where $\tau_{\theta} \mathfrak{m}(\xi) := \mathfrak{m}(\xi - \theta)$ for $\xi \in \mathcal{G}^*$.

For every $k \in \mathbb{Z}_+$, we denote by $(\mathcal{G}^*)^k$ the direct product of k copies of groups \mathcal{G}^* , and denote by m a smooth tempered function from $(\mathcal{G}^*)^k$ to \mathbb{C} . We define the multilinear Fourier multiplier operator $B_{\mathcal{G}}[m]: (\mathbf{S}(\mathcal{G}))^k \to \mathbf{S}(\mathcal{G})$ by the formula

$$B_{\mathcal{G}}[m](f_1, \dots, f_k)(x) := \int_{(\mathcal{G}^*)^k} m(\xi) \Big(\prod_{i \in [k]} \mathcal{F}_{\mathcal{G}} f_i(\xi_i) \Big) e(-x \cdot (\xi_1 + \dots + \xi_k)) \ d\nu_{(\mathcal{G}^*)^k}(\xi), \quad (2.9)$$

where $x \in \mathcal{G}$, $\xi = (\xi_1, \dots, \xi_k)$ and $\nu_{(\mathcal{G}^*)^k}(\xi) = \prod_{i \in [k]} \nu_{\mathcal{G}^*}(\xi_i)$. We refer to m as the *symbol* of $B_{\mathcal{G}}[m]$.

2.4. Terminology from the classical circle method. We fix some notation and terminology from the classical circle method. Define the height $h(\alpha) \in 2^{\mathbb{N}}$ of an arithmetic frequency $\alpha = \frac{b}{a} \mod 1 \in \mathbb{Q}/\mathbb{Z}$ by the formula

$$h\left(\frac{b}{q} \bmod 1\right) := \inf\{2^l : 2^l \ge q, l \in \mathbb{N}\} = 2^{\log q} \sim q$$

whenever $q \in \mathbb{Z}_+$ and $b \in [q]^{\times}$. For any $l \in \mathbb{N}$, $k \in \mathbb{Z}$, we define the arithmetic frequency sets

$$(\mathbb{Q}/\mathbb{Z})_{\leq l} := \{ \alpha \in \mathbb{Q}/\mathbb{Z} : h(\alpha) \leq 2^{l} \}, \qquad (\mathbb{Q}/\mathbb{Z})_{l} := (\mathbb{Q}/\mathbb{Z})_{\leq l} \setminus (\mathbb{Q}/\mathbb{Z})_{\leq l-1}, \tag{2.10}$$

(with the convention that $(\mathbb{Q}/\mathbb{Z})_{\leq -1} = \emptyset$) and the continuous frequency sets

$$\mathbb{R}_{\leq k} := [-2^k, 2^k].$$

Then we define the major arcs

$$\mathfrak{M}_{\langle l, \langle k} := \pi(\mathbb{R}_{\langle k} \times (\mathbb{Q}/\mathbb{Z})_{\langle l})$$
(2.11)

with the map π defined by (B.1) in Appendix B. Thus, $\mathfrak{M}_{\leq l, \leq k}$ consists of all elements of \mathbb{T} of the form $\frac{b}{q} + \theta$ mod 1 for some $q \in [2^l]$, $b \in [q]^{\times}$, and $\theta \in [-2^k, 2^k]$. These major arcs are the natural focus for our Fourier-analytic manipulations. Note that the height in this paper aligns with the naive height from [31]. Since we will apply the Ionescu-Wainger multiplier theorem for canonical fractions (see Section 2.5), there is no need to introduce the height from [31].

Let $(l, k) \in \mathbb{N} \times \mathbb{Z}$, and recall definitions (2.8), (2.1). We introduce the Ionescu-Wainger projections by setting

$$\Pi_{\leq l, \leq k} f(x) := T_{\mathbb{Z}}^{\mathcal{R}_{\leq 2^l}} [\eta_{\leq k}] f(x) \quad \text{for} \quad x \in \mathbb{Z},$$
(2.12)

where the set $\mathcal{R}_{\leq 2^l}$ is defined by

$$\mathcal{R}_{\leq 2^l} := \{ b/q \in \mathbb{Q} \cap \mathbb{T} : q \in [2^l] \text{ and } (b,q) = 1 \}.$$
 (2.13)

Remark 1. One can easily see that the set $\mathcal{R}_{\leq 2^l}$ plays the same role as the above $(\mathbb{Q}/\mathbb{Z})_{\leq l}$. In fact, the notation (2.10) is employed to streamline the analysis of adelic groups and related abstract groups in Section 6, whereas (2.13) is used mainly for convenience in the notation of Fourier multiplier operators like (2.12).

2.5. Ionescu-Wainger multiplier theorem for canonical fractions. Invoking the definitions of $T_{\mathcal{G}}^{\mathfrak{Q}}[\mathfrak{m}]$ from (2.8) and $T_{\mathcal{G}}[\mathfrak{m}]$ from (2.7), we introduce below the Ionescu-Wainger multiplier theorem for canonical fractions proved in the recent breakthrough work by Kosz-Mirek-Peluse-Wright [30].

Theorem 2.1. Let $q \in [p'_0, p_0]$ for some $p_0 \in 2\mathbb{Z}_+$. Then there exists a constant $C_{p_0} \in \mathbb{R}_+$ such that for every integer $N \geq 100$ the following is true. Assume that

$$0 < \vartheta \le (2p_0 N^{p_0})^{-1}$$

and let $\mathfrak{m}: \mathbb{R}^d \to L(H_1, H_2)$ be a measurable function supported on $[-\vartheta, \vartheta]$, whose values are bounded linear operators between two separable Hilbert spaces H_1 and H_2 . Then

$$||T_{\mathbb{Z}}^{\mathcal{R}_{\leq N}}[\mathfrak{m}]||_{\ell^{q}(\mathbb{Z};H_{1})\to\ell^{q}(\mathbb{Z};H_{2})} \lesssim_{p_{0}} 2^{\mathbf{C}_{p_{0}}(N)} ||T_{\mathbb{R}}[\mathfrak{m}]||_{L^{p_{0}}(\mathbb{R};H_{1})\to L^{p_{0}}(\mathbb{R};H_{2})}, \tag{2.14}$$

where the power $\mathbf{C}_{p_0}(N)$ is given by

$$\mathbf{C}_{p_0}(N) := C_{p_0} \log N \frac{\log \log \log N}{\log \log N}.$$
 (2.15)

Indeed, [30] obtained a general dimensional Ionescu-Wainger multiplier theorem for canonical fractions. Here we focus on the one-dimensional result presented in Theorem 2.1, as it suffices for our proof. In particular, we will use (2.14) with $H_1 = H_2 = \mathbb{C}$, and $H_1 = H_2 = \ell^2$. The lower bound for N is imposed to ensure the logarithm in (2.15) remains well-defined, though it is not essential to our proof.

2.6. Approximant to the von Mangoldt function. Let μ be the Möbius function, defined by $\mu(n) = (-1)^k$ if n is the product of k distinct primes and by $\mu(n) = 0$ if n is divisible by the square of a prime; see [28, Chapter 1]. For $\omega \geq 1$, we introduce the Cramér approximant $\Lambda_{\text{Cramér},\omega}$ given by

$$\Lambda_{\operatorname{Cram\acute{e}r},\omega}(n) = \frac{W}{\varphi(W)} \mathbb{1}_{(n,W)=1}, \quad \text{where} \quad W = \prod_{n \le \omega} p, \tag{2.16}$$

and the Heath-Brown approximant $\Lambda_{\mathrm{HB},\omega}$ given by

$$\Lambda_{\mathrm{HB},\omega}(n) = \sum_{q < \omega} \frac{\mu(q)}{\varphi(q)} c_q(n), \quad \text{where} \quad c_q(n) = \sum_{r \in [q]^{\times}} e(rn/q). \tag{2.17}$$

(By convention, $\Lambda_{\operatorname{Cram\acute{e}r},\omega} = \Lambda_{\operatorname{HB},\omega} = 0$ if $\omega = 1$.) For any $k \in \mathbb{Z}_+$ and $N \geq 1$, the Cramér approximant $\Lambda_{\operatorname{Cram\acute{e}r},\omega}$ and the von Mangoldt function Λ satisfy

$$0 \le \Lambda_{\operatorname{Cram\acute{e}r},\omega} \lesssim \langle \operatorname{Log} \omega \rangle \quad \text{and} \quad \mathbb{E}_{n \in [N]} \Lambda(n) \lesssim 1,$$
 (2.18)

while the Heath-Brown approximant $\Lambda_{\mathrm{HB},\omega}$ obeys the moment bounds

$$\mathbb{E}_{n \in [N]} |\Lambda_{\mathrm{HB},\omega}(n)|^k \lesssim_k \langle \mathrm{Log}\,\omega \rangle^{2^k + k}. \tag{2.19}$$

Indeed, (2.18) is a direct result of $\prod_{p \leq \omega} p/(p-1) \lesssim \langle \text{Log } \omega \rangle$ and the definition of Λ ; for the proof of (2.19), we refer to [32, Lemma 4.6].

In the proof of our main result, more precise estimates related to the "little" Gowers uniformity norm for the above approximants are needed. We first introduce the definition of this norm. For $s \in \mathbb{N}$, we define the little Gowers norms $u^{s+1}[N]$ of a function $f : \mathbb{Z} \to \mathbb{C}$ by

$$||f||_{u^{s+1}[N]} := \sup_{\deg P \le s} |\mathbb{E}_{n \in [N]} f(n) e(P(n))|,$$
 (2.20)

where P ranges over all real-coefficient polynomials of degree at most s. For more details on the little Gowers norms and the related Gowers norms, we refer to [53, 22, 23]. We now list some useful estimates involving the above approximants.

Lemma 2.2. (i) Let $N \ge 100$ and $1 \le z \le \exp(\text{Log}^{1/10}N)$. Then there exists a positive constant c independent of N such that

$$\mathbb{E}_{n \in [N]} \Lambda_{\operatorname{Cram\acute{e}r}, z}(n) = 1 + O(\exp(-c \operatorname{Log}^{4/5} N)); \tag{2.21}$$

moreover, if $1 \le q \le z$, $b \in [q]$ and I is an interval of length N,

$$\mathbb{E}_{n \in I} \Lambda_{\operatorname{Cram\acute{e}r}, z}(n) \mathbb{1}_{n \equiv b \pmod{q}} = \frac{\mathbb{1}_{(b,q)=1}}{\varphi(q)} + O(\exp(-c\operatorname{Log}^{4/5}N)). \tag{2.22}$$

(ii) Let $N \ge 100$, $d \in \mathbb{Z}_+$ and $1 \le z, w \le \exp(\operatorname{Log}^{1/10} N)$. Then there exists a positive constant c_d , depending only on d, such that

$$\|\Lambda_{\operatorname{Cram\acute{e}r},w} - \Lambda_{\operatorname{Cram\acute{e}r},z}\|_{u^{d+1}(I)} \lesssim_d w^{-c_d} + z^{-c_d}$$
(2.23)

for any interval I of length N.

(iii) Let $N \ge 1$, $d \in \mathbb{Z}_+$ and $1 \le w, Q, Q_1, Q_2 \le \exp(\operatorname{Log}^{1/20} N)$. Then there exists a positive constant c'_d , depending only on d, such that

$$\|\Lambda_{\text{Cram\'er},w} - \Lambda_{\text{HB},Q}\|_{u^{d+1}(I)} \lesssim_d w^{-c'_d} + Q^{-c'_d} \quad \text{and}$$
 (2.24)

$$\|\Lambda_{\mathrm{HB},Q_1} - \Lambda_{\mathrm{HB},Q_2}\|_{u^{d+1}(I)} \lesssim_d Q_1^{-c'_d} + Q_2^{-c'_d}$$
(2.25)

for any interval I of length N.

For the proofs of (2.21)-(2.22), (2.23), and (2.24)-(2.25), we refer to Corollary 4.4, Lemma 4.5, and Proposition 4.7 in [32] respectively. In particular, the fundamental lemma of sieve theory (see [28]) and some important arguments in [52, 36] were used in these proofs. Here, the restriction $N \geq 100$ in both (i) and (ii) does not affect our subsequent proof, as we will only focus on the case where $N \geq C$ with sufficiently large C > 0.

2.7. Generalized von Neumann theorem by Teräväinen. Due to the potential presence of Siegel zeros (see [32, Subsections 1.4 and 6.2] for the details) when analyzing the symbol of the averages weighted by the von Mangoldt function Λ , we employ the following generalized von Neumann theorem to reduce the matter to bounding the Cramér-weighted averages.

Lemma 2.3. Let $d, k \in \mathbb{Z}_+$, $C \geq 1$ and let $\mathcal{P} = \{P_1, \ldots, P_k\}$ be a family of polynomials with integer coefficients and distinct degrees, where $\max\{\deg P_i : i \in [k]\} = d$. Let $N \geq 1$, $f_0, \ldots, f_k : \mathbb{Z} \to \mathbb{C}$ be 1-bounded functions supported on $[\pm CN^d]$, and let $\theta : [N] \to \mathbb{C}$ be a 1-bounded function. Then there exists some $1 \leq K \lesssim_d 1$ such that

$$\left| \frac{1}{N^{d+1}} \sum_{m \in \mathbb{Z}} \sum_{n \in [N]} \theta(n) f_0(m) f_1(m + P_1(n)) \cdots f_k(m + P_k(n)) \right| \lesssim_{C, \mathcal{P}} (N^{-1} + \|\theta\|_{u^{d+1}[N]})^{1/K},$$

where $u^{d+1}[N]$ is defined by (2.20) with s = d.

Teräväinen [53] proved this lemma using Peluse's inverse theorem [47, Theorem 3.3] and some crucial reduction arguments. Additionally, the condition $|\theta| \leq 1$ explains why, in the proof of Proposition 3.4 below, we cannot directly reduce the matter to estimating the averages weighted by the function $\Lambda_{\text{HB},\omega}$ (which does not have a good upper bound).

3. Multilinear Weyl inequality with Cramér weight

In this section, we shall prove a multilinear Weyl inequality⁵ (minor arcs estimates) for the Cramér-weighted averages, which asserts that the Cramér-weighted averages are negligible when the Fourier transform of f_j , for at least one $j \in [k]$, vanishes on appropriate major arcs. Consider a polynomial mapping

$$\mathcal{P} := (P_1, \dots, P_k) : \mathbb{Z} \to \mathbb{Z}^k, \tag{3.1}$$

where P_1, \ldots, P_k are polynomials with integer coefficients and distinct degrees such that

$$1 < d_1 := \deg P_1 < \dots < d_k := \deg P_k. \tag{3.2}$$

⁵We use the terminology "multilinear Weyl inequality" throughout this paper since this inequality can be regarded as a variant of the linear Weyl inequality. Moreover, this terminology has also been employed in [30] to investigate the pointwise convergence of the general unweighted averages (1.3).

Throughout this paper, we fix a large constant C_0 (say, $C_0 = 100$), and consider the Cramér approximant (see (2.16) for the definition)

$$\Lambda_N(n) := \Lambda_{\operatorname{Cram\acute{e}r.exp}(\operatorname{Log}^{1/C_0} N)}(n). \tag{3.3}$$

We now formulate below the weighted multilinear Weyl inequality.

Theorem 3.1. Let $N \geq 1$, $l \in \mathbb{N}$, $k \in \mathbb{Z}_+$, and let \mathcal{P} be a polynomial mapping satisfying (3.1) and (3.2). Let $1 < q_1, \ldots, q_k < \infty$ be exponents such that $\frac{1}{q_1} + \cdots + \frac{1}{q_k} = \frac{1}{q} \leq 1$. There exists a small $c \in (0,1)$, depending on $k, \mathcal{P}, q, q_1, \ldots, q_k$, such that the following holds. Let $f_i \in \ell^{q_i}(\mathbb{Z})$ for each $i \in [k]$, and fix $j \in [k]$. If $f_j \in \ell^{q_j}(\mathbb{Z}) \cap \ell^2(\mathbb{Z})$, and $\mathcal{F}_{\mathbb{Z}} f_j$ vanishes on the major arcs $\mathfrak{M}_{< l, < -d_j \operatorname{Log} N + d_j l}$ given by (2.11),

$$\|\tilde{A}_{N,\Lambda_N}^{\mathcal{P}}(f_1,\ldots,f_k)\|_{\ell^{q_1}(\mathbb{Z})} \le c^{-1} (2^{-cl} + \exp(-c \log^{1/C_0} N)) \|f_1\|_{\ell^{q_1}(\mathbb{Z})} \cdots \|f_k\|_{\ell^{q_k}(\mathbb{Z})}.$$
(3.4)

The same conclusion holds for $A_{N,\Lambda_N}^{\mathcal{P}}$ in place of $\tilde{A}_{N,\Lambda_N}^{\mathcal{P}}$.

- Remark 2. (1) This bound $\exp(-c \operatorname{Log}^{1/C_0} N)$ is determined by the lower bound in (3.22) of Theorem 3.3. Furthermore, the upper bound $2^{-cl} + \exp(-c \operatorname{Log}^{1/C_0} N)$ does not meet the polynomial-type requirement $2^{-cl} + N^{-c}$ of the method in [30].
- (2) It suffices to prove Theorem 3.1 under this assumption $2^l \lesssim \exp(\log^{1/C_0} N)$, since for larger values of l, the support condition becomes stronger, and the conclusion (3.4) is essentially unchanged. Moreover, we may further assume that l is sufficiently large, since the claim follows from Minkowski's inequality, (2.21) and Hölder's inequality otherwise.
- 3.1. **Inverse theorems.** We use the inner product $\langle f,g\rangle:=\sum_{x\in\mathbb{Z}}f(x)g(x)$ on $\mathbf{S}(\mathbb{Z})$. (In the following multilinear analysis, inserting a complex conjugation into the inner product may not provide any advantage.) Let $k\in\mathbb{Z}_+$, $h,f_1,\ldots,f_k\in\mathbf{S}(\mathbb{Z})$, $N\geq 1$, and $j\in[k]$. Let $w:\mathbb{Z}_+\to\mathbb{C}$ be a weight function. Denote $\tilde{A}_{N,w}^{\mathcal{P}}:=\tilde{A}_{N,w;\mathbb{Z}}^{\mathcal{P}}$ and $A_{N,w}^{\mathcal{P}}:=A_{N,w;\mathbb{Z}}^{\mathcal{P}}$, and observe the identity

$$\langle \tilde{A}_{N,w}^{\mathcal{P}}(f_1,\ldots,f_k), h \rangle = \langle \tilde{A}_{N,w}^{\mathcal{P},*j}(f_1,\ldots,f_{j-1},h,f_{j+1},\ldots,f_k), f_j \rangle, \tag{3.5}$$

where the dual operator $\tilde{A}_{N,w}^{\mathcal{P}*j}$ is given by

$$\tilde{A}_{N,w}^{\mathcal{P},*j}(g_1,\ldots,g_k)(x) := \frac{1}{\lfloor N \rfloor} \sum_{n \in J_N} w(n) \prod_{i \in [k]} g_i \left(x - \mathbb{1}_{i \neq j} P_i(n) + P_j(n) \right)$$
(3.6)

with $J_N = [N] \setminus [N/2]$. A similar identity also holds for $A_{N,w}^{\mathcal{P}}$ and its dual operator $A_{N,w}^{\mathcal{P},*j}$ for each $j \in [k]$. Additionally, we shall often abbreviate

$$\mathfrak{A}_{N,w} := \mathfrak{A}_{N,w}^{\mathcal{P}}, \quad \mathfrak{A}_{N,w}^{*j} := \mathfrak{A}_{N,w}^{\mathcal{P},*j} \quad \text{whenever} \quad \mathfrak{A} \in \{A, \tilde{A}\}.$$

3.1.1. Unweighted inverse theorem. We first list an important inverse theorem for the unweighted averages. One can see similar theorems in [31, 30].

Theorem 3.2. Let $N \geq 1$, $0 < \delta \leq 1$, $k \in \mathbb{Z}_+$, and let \mathcal{P} be a polynomial mapping satisfying (3.1) and (3.2). Let N_* be a quantity with $N_* \sim N^{d_k}$, and fix $j \in [k]$. Let $h, f_1, \ldots, f_k \in \mathbf{S}(\mathbb{Z})$ be 1-bounded functions supported on $[\pm N_*]$, obeying the lower bound

$$\left|\left\langle \tilde{A}_{N,1}(f_1,\ldots,f_k),h\right\rangle\right| \ge \delta N^{d_k}.$$
 (3.7)

Then there exists a function $F_j \in \ell^2(\mathbb{Z})$ with

$$||F_j||_{\ell^{\infty}(\mathbb{Z})} \lesssim 1, \quad ||F_j||_{\ell^1(\mathbb{Z})} \lesssim N^{d_k} \tag{3.8}$$

and with $\mathcal{F}_{\mathbb{Z}}F_j$ supported in the $O(\delta^{-O(1)}/N^{d_j})$ -neighborhood of some $\alpha_j \in \mathbb{Q}/\mathbb{Z}$ of height $O(\delta^{-O(1)})$ such that

$$|\langle f_j, F_j \rangle| \gtrsim \delta^{O(1)} N^{d_k}$$
.

The same conclusion holds for the operator $A_{N,1}$ in place of $\tilde{A}_{N,1}$ in (3.7).

Peluse's inverse theorem (see [47, Theorem 3.3] or [31, Theorem 6.3]) will play a crucial role in our proof. Before proceeding, some supplementary remarks are necessary to clarify the ideas. Firstly, it suffices to consider δ to be sufficiently small. Secondly, any polynomial in \mathcal{P} , whether or not it has a non-zero constant term, does not affect the above conclusions, because the translation of a function on \mathbb{Z} does not change its Fourier support.

Proof of Theorem 3.2. Note that we can simply take $F_j = \tilde{A}_{N,1}^{*j}(f_1, \ldots, f_{j-1}, h, f_{j+1}, \ldots, f_k)$ defined by (3.6) whenever $N \lesssim \delta^{-O(1)}$. Thus, we only consider the case

$$N \ge \mathcal{C}\delta^{-\mathcal{C}} \tag{3.9}$$

with sufficiently large constant C. By repeating the arguments yielding [31, Theorem 6.3 and Proposition 6.6], we can infer that

Theorem 3.2 for the case
$$j = 1$$
 holds. (3.10)

We next prove Theorem 3.2 for the case i = 2.

Without loss of generality, we may assume that δ is sufficiently small. Define

$$G_{k,N}(x) := \tilde{A}_{N,1}^{*1}(h, f_2, \dots, f_k)(x).$$
 (3.11)

Obviously, $G_{k,N}$ is 1-bounded function and supported on $[\pm N'_*]$ with $N_* \leq N'_* \sim N^{d_k}$. By (3.7), (3.5), (3.11) and Hölder's inequality,

$$N^{d_k} |\langle G_{k,N}, \tilde{A}_{N,1}^{*1}(h, f_2, \dots, f_k) \rangle| \gtrsim |\langle G_{k,N}, f_1 \rangle|^2 \geq \delta^2 N^{2d_k},$$

which, with (3.5), yields

$$|\langle \tilde{A}_{N,1}(G_{k,N}, f_2, \dots, f_k), h \rangle| \gtrsim \delta^2 N^{d_k}.$$

By Theorem 3.2 (for the case j=1) with f_1 replaced by $G_{k,N}$ (since $G_{k,N}, f_2, \ldots, f_k$ are 1-bounded functions supported on $[\pm N'_*]$), there exist a constant $C_* > 0$ and a function $F_1 \in \ell^2(\mathbb{Z})$ with (3.8) and with $\mathcal{F}_{\mathbb{Z}}F_1$ supported in the δ^{-C_*}/N^{d_1} -neighborhood of some $\alpha = \frac{a}{b} \mod 1$ in \mathbb{Q}/\mathbb{Z} obeying $b = O(\delta^{-C_*})$, such that $|\langle G_{k,N}, F_1 \rangle| \gtrsim \delta^{C_*}N^{d_k}$, which, with (3.5) and (3.11), gives

$$|\langle \tilde{A}_{N,1}(F_1, f_2, \dots, f_k), h \rangle| \gtrsim \delta^{C_*} N^{d_k}. \tag{3.12}$$

We will reduce the average $\tilde{A}_{N,1}$ concerning k functions F_1, f_2, \ldots, f_k in inequality (3.12) to some average only involving k-1 functions f_2, \ldots, f_k . For $m \in J_N$, we define

$$F_{1,\alpha,m}(x) := F_1(x - P_1(m))e(\frac{a}{b}(x - P_1(m)))$$

$$= \int_{\mathbb{T}} e(-(\xi - \frac{a}{b}) \cdot (x - P_1(m)))\mathcal{F}_{\mathbb{Z}}F_1(\xi)d\xi, \quad x \in \mathbb{Z},$$
(3.13)

which, with (3.9) and the support condition for $\mathcal{F}_{\mathbb{Z}}F_1$, gives that for $C_1 \in [2C_*, \mathcal{C}/2]$,

$$F_{1,\alpha,m}(x) = \int_{\mathbb{R}} e\left(-(\xi - \frac{a}{b}) \cdot (x - P_1(m))\right) \eta(\delta^{C_1} N^{d_1}(\xi - \frac{a}{b})) \mathcal{F}_{\mathbb{Z}} F_1(\xi) d\xi, \quad x \in \mathbb{Z}.$$
 (3.14)

For any pair $n_1, n_2 \in J_N$ with $|n_1 - n_2| \leq \delta^{2C_1} N$, we **claim** the local constant property

$$F_{1,\alpha,n_1}(x) = F_{1,\alpha,n_2}(x) + O(\delta^{C_1}), \tag{3.15}$$

whose proof is postponed until the end of this subsection. Splitting the interval J_N into subintervals of length $\delta^{2C_1}N$, we deduce by the pigeonhole principle and (3.12) that there is at least one interval of length $\delta^{2C_1}N$, denoted by $I := \mathbb{Z} \cap [c_I, c_I + \delta^{2C_1}N]$ $(c_I \in J_N)$, such that

$$|\langle \tilde{A}_{I,1}(F_1, f_2, \dots, f_k), h \rangle| \gtrsim \delta^{C_*} N^{d_k}, \tag{3.16}$$

where the operator $\tilde{A}_{I,1}$ is defined by

$$\tilde{A}_{I,1}(F_1, f_2, \dots, f_k)(x) := \mathbb{E}_{n \in I} F_1(x - P_1(n)) f_2(x - P_2(n)) \cdots f_k(x - P_k(n)).$$

Using (3.13) and (3.15) with $(n_1, n_2) = (n, c_I)$, we have

$$\tilde{A}_{I,1}(F_1, f_2, \dots, f_k)(x) = F_{1,\alpha,c_I}(x)e(-\frac{a}{b}x)A_I^{(1)}(f_2, \dots, f_k)(x) + E(x)$$

where the error $E(x) = O(\delta^{C_1})$ and the operator $A_I^{(1)}$ is given by

$$A_I^{(1)}(f_2,\dots,f_k) := \mathbb{E}_{n\in I}e(\frac{a}{b}P_1(n))f_2(x-P_2(n))\dots f_k(x-P_k(n)).$$

Note that $|\langle E, h \rangle| \lesssim \delta^{C_1} N^{d_k}$ and C_1 is large enough. Thus (3.16) and (3.13) can yield

$$|\langle A_I^{(1)}(f_2,\cdots,f_k),h_1\rangle|\gtrsim \delta^{C_*}N^{d_k}$$

with $h_1(x) := h(x)F_{1,\alpha,c_I}(x)e(-ax/b) = h(x)F_1(x-P_1(c_I))e(-aP_1(c_I)/b)$. Using (3.8) with j=1, we deduce that h_1/C_2 (for some sufficiently large C_2) is 1-bounded functions supported on $[\pm N_*]$. By the triangle inequality, we further infer

$$|\langle A_{I,N'}^{(1)}(f_2,\cdots,f_k),h_1\rangle| \gtrsim \delta^{C_*}N^{d_k},$$
 (3.17)

for either $N' = c_I$ or $N' = c_I + \delta^{2C_1}N$, where

$$A_{I,N'}^{(1)}(f_2,\cdots,f_k)(x) := \mathbb{E}_{n\in[N']}e(\frac{a}{b}P_1(n))f_2(x-P_2(n))\cdots f_k(x-P_k(n)).$$

We split the issue into two cases: (1) a/b = 0 and (2) $1 \le a < b = O(\delta^{-C_*})$. For the first case $(e(aP_1(n)/b) = 1)$, from (3.10) we can end the proof of Theorem 3.2 (for δ^{C_*} in place of δ). Next, we consider the second case. We may assume that (3.17) holds for $N' = c_I$ since the situation that (3.17) holds for $N' = c_I + \delta^{2C_1}N$ can be treated by the same way. By writing n = mb + r with $m \in \mathbb{Z}$ and $r \in [b]$ (i.e., the congruence-based factorization),

$$A_{L,r}^{(1)}(f_2,\cdots,f_k) = \mathbb{E}_{r\in[b]} \ e(aP_1(r)/b) \ A_{L,r}^{(2)}(\tau_{P_2(r)}f_2,\cdots,\tau_{P_k(r)}f_k) + O(N^{-1/2}),$$

where $\tau_{P_i(r)} f_i := f_i(\cdot - P_i(r))$ for $i = 2, \dots, k$, and the operator $A_{I,r}^{(2)}$ is given by

$$A_{I,r}^{(2)}(g_2,\cdots,g_k) := \mathbb{E}_{m\in[c_I/b]} g_2(x-\tilde{P}_{2,b,r}(m))\cdots g_k(x-\tilde{P}_{k,b,r}(m))$$

with $\tilde{P}_{i,b,r}(m) = P_i(mb+r) - P_i(r)$ for i = 2..., k. Using the pigeonhole principle again and invoking (3.9), we deduce that there is $r_0 \in [b]$ such that

$$|\langle A_{I,r_0}^{(2)}(\tau_{P_2(r_0)}f_2,\cdots,\tau_{P_k(r_0)}f_k),h_1\rangle|\gtrsim \delta^{C_*}N^{d_k}.$$

Since $b = O(\delta^{-C_*})$, these polynomials $\{\tilde{P}_{i,b,r}\}_{i=2,\dots,k}$ have $(C_{\mathcal{P}}, \delta^{-O(1)})$ -coefficients (see [47, Definition 3.1] for the definition) for some constant $C_{\mathcal{P}}$. Thus we can use [47, Theorem 3.3] for polynomials $\{\tilde{P}_{i,b,r}\}_{i=2,\dots,k}$ to achieve an estimate for $\tau_{P_2(r_0)}f_2$ (i.e., the inequality in [31, Theorem 6.3 (ii)] with f replaced by $\tau_{P_2(r_0)}f_2$), which, together with the arguments leading to [31, Proposition 6.6], yields the conclusion in Theorem 3.2 for the case j=2 (since the translation of a function on $\mathbb Z$ does not change its Fourier support).

Similarly, repeating the above process, we can achieve Theorem 3.2 for the case j = 3, ..., k in order whenever $k \geq 3$. This completes the proof of Theorem 3.2 by accepting (3.15). It remains to show (3.15). By (3.14) and Taylor's expansion

$$e\left(-(\xi-\frac{a}{b})\cdot(P_1(n_2)-P_1(n_1))\right)=1+\sum_{\nu=1}^{\infty}\frac{\left(2\pi i\right)^{\nu}}{\nu!}(\xi-\frac{a}{b})^{\nu}(P_1(n_2)-P_1(n_1))^{\nu},$$

we write function F_{1,α,n_1} as

$$\int_{\mathbb{R}} e\left(-(\xi - \frac{a}{b}) \cdot (x - P_1(n_1))\right) \eta(\delta^{C_1} N^{d_1}(\xi - \frac{a}{b})) \mathcal{F}_{\mathbb{Z}} F_1(\xi) d\xi$$

$$= F_{1,\alpha,n_2}(x) + \sum_{\nu=1}^{\infty} \frac{(2\pi i)^{\nu}}{\nu!} \delta^{-C_1 \nu} N^{-d_1 \nu} (P_1(n_2) - P_1(n_1))^{\nu} T_{\nu,\alpha} F_1(x - P_1(n_2)), \tag{3.18}$$

where the function $T_{\nu,\alpha}F_1$ is given by

$$T_{\nu,\alpha}F_1(x) := \int_{\mathbb{R}} e\left(-(\xi - \frac{a}{b}) \cdot x\right) \tilde{\eta}_{\nu}(\delta^{C_1}N^{d_1}(\xi - \frac{a}{b})) \mathcal{F}_{\mathbb{Z}}F_1(\xi) d\xi$$
$$= \int_{\mathbb{R}} e\left(-\xi \cdot x\right) \tilde{\eta}_{\nu}(\delta^{C_1}N^{d_1}\xi) \mathcal{F}_{\mathbb{Z}}F_1(\xi + \frac{a}{b}) d\xi \quad (x \in \mathbb{Z})$$

with $\tilde{\eta}_{\nu}(\xi) = \xi^{\nu} \eta(\xi)$ satisfying a direct bound

$$|\mathcal{F}_{\mathbb{R}}^{-1}\tilde{\eta}_{\nu}(z)| \lesssim C_{\circ}^{\nu} \langle z \rangle^{-2} \tag{3.19}$$

for some absolute constant $C_{\circ} > 0$. Notice that the mean value theorem gives

$$P_1(n_1) - P_1(n_2) = O(N^{d_1 - 1}|n_1 - n_2|) = O(\delta^{2C_1}N^{d_1}).$$
(3.20)

Substituting (3.20) into (3.18), we reduce the proof of (3.15) to showing

$$||T_{\nu,\alpha}F_1||_{\ell^{\infty}(\mathbb{Z})} \le C_{\circ}^{\nu}||F_1||_{\ell^{\infty}(\mathbb{Z})}.$$
 (3.21)

By expanding the definition of $\mathcal{F}_{\mathbb{Z}}F_1(\xi)$, we can write

$$T_{\nu,\alpha}\left(e(-\frac{a}{b}\cdot)F_1\right)(x) = \int_{\mathbb{R}} e\left(-\xi\cdot x\right)\tilde{\eta}_{\nu}(\delta^{C_1}N^{d_1}\xi)\mathcal{F}_{\mathbb{Z}}F_1(\xi)d\xi$$
$$= \sum_{y\in\mathbb{Z}} F_1(y)\delta^{-C_1}N^{-d_1}(\mathcal{F}_{\mathbb{R}}^{-1}\tilde{\eta}_{\nu})(\delta^{-C_1}N^{-d_1}(y-x)),$$

which, combined (3.9) and (3.19), yields $|T_{\nu,\alpha}(e(-\frac{a}{b}\cdot)F_1)| \leq C_{\circ}^{\nu}M_{DHL}F_1$ (M_{DHL} denotes the discrete Hardy-Littlewood maximal operator). Finally, we have $||T_{\nu,\alpha}(e(-\frac{a}{b}\cdot)F_1)||_{\ell^{\infty}(\mathbb{Z})} \leq C_{\circ}^{\nu}||F_1||_{\ell^{\infty}(\mathbb{Z})}$, which leads to (3.21) immediately.

3.1.2. Weighted inverse theorem. This subsection presents the weighted extension of Theorem 3.2. A crucial distinction from the unweighted case is the necessity to impose a lower bound constraint on the parameter δ in the weighted setting.

Theorem 3.3. Let N, k, \mathcal{P} and N_* be given as in Theorem 3.2. Fix $j \in [k]$ and assume

$$\exp(-c'\log^{1/C_0}N) \le \delta \le 1. \tag{3.22}$$

with sufficiently small $c' = c'(k, d_k) > 0$. Let $h, f_1, \ldots, f_k \in \mathbf{S}(\mathbb{Z})$ be 1-bounded functions supported on $[\pm N_*]$, obeying the lower bound

$$\left| \left\langle \tilde{A}_{N,\Lambda_N}(f_1,\dots,f_k), h \right\rangle \right| \ge \delta N^{d_k}. \tag{3.23}$$

Then the conclusions of Theorem 3.2 hold.

As in the proof of Theorem 3.2, it suffices to prove Theorem 3.3 under the assumption that (3.9) holds. We need the following proposition, which crucially depends on (3.22).

Proposition 3.4. Under the hypotheses and notation of Theorem 3.3 with additional (3.9), there exists a positive $\omega_{\circ} = O(\delta^{-O(1)})$ such that

$$|\langle \tilde{A}_{N,\Lambda_{\mathrm{HB},\omega_o}}(f_1,\cdots,f_k),h\rangle| \gtrsim \delta N^{d_k},$$
 (3.24)

where the weight Λ_{HB,ω_0} is a Heath-Brown approximant given by (2.17) with $\omega=\omega_0$.

Proof of Theorem 3.3 (accepting Proposition 3.4). Invoking the definition of Λ_{HB,ω_o} (2.17), we deduce from (3.24) that there exist $q \in |\omega_o|$ and $r \in [q]^\times$ such that

$$|\langle \tilde{A}_{N,e(r,q)}(f_1,\cdots,f_k),h\rangle| \gtrsim \delta^{O(1)} N^{d_k}. \tag{3.25}$$

We begin by categorizing the problem into two cases, depending on whether the polynomial P_1 in \mathcal{P} is linear. We first prove the linear case, then deal with the nonlinear case.

Case 1. For the linear case, we assume $P_1(n) = an + b$ for some $a, b \in \mathbb{Z}$ with $a \neq 0$. Write

$$e(\frac{rn}{q}) = e\left(-\frac{r}{aq}(x-an-b)\right)e\left(\frac{r}{aq}(x-b)\right), \quad \tilde{f}_1 := e(-\frac{r}{aq}\cdot)f_1, \quad \tilde{h} := e(\frac{r}{aq}(\cdot-b))h.$$

Then (3.25) gives

$$|\langle \tilde{A}_N(\tilde{f}_1, f_2, \dots, f_k), \tilde{h} \rangle| \gtrsim \delta^{O(1)} N^{d_k}.$$

Note that $\tilde{f}_1, \tilde{h} \in \mathbf{S}(\mathbb{Z})$ are also 1-bounded functions supported on $[\pm N_*]$. Hence, Theorem 3.2 gives that, for $j \in [k]$, there exists a function $G_j \in \ell^2(\mathbb{Z})$ satisfying desired ℓ^q bounds as in (3.8), and obeying that $\mathcal{F}_{\mathbb{Z}}G_j$ is supported in the $O(\delta^{-O(1)}/N^{d_j})$ -neighborhood of some $\alpha_j \in \mathbb{Q}/\mathbb{Z}$ of height $O(\delta^{-O(1)})$ such that

$$|\langle \mathbb{1}_{j=1}\tilde{f}_1 + \mathbb{1}_{j\neq 1}f_j, G_j\rangle| \gtrsim \delta^{O(1)}N^{d_k}$$

From this bound we can achieve the goal by setting

$$F_j := \mathbb{1}_{j=1} e(-\frac{r}{aa}\cdot)G_1 + \mathbb{1}_{j\neq 1}G_j.$$

As a result, for any $k \in \mathbb{N}$ and any $N_* \sim N^{d_k}$, we obtain the conclusions in Theorem 3.3 for the case where P_1 is linear.

Case 2. We consider the case where $\deg P_1 \geq 2$. Obviously, it seems difficult to directly use the above approach treating the linear case. Here we use a new reduction argument. Using (3.25), Fubini's theorem, and the change of variables $x \to x + n$, we have

$$\left| \mathbb{E}_{n \in J_N} e(rn/q) \left(\sum_{x \in \mathbb{Z}} f_1(x - \tilde{P}_1(n)) \cdots f_k(x - \tilde{P}_k(n)) h(x + n) \right) \right| \gtrsim \delta^{O(1)} N^{d_k}$$
 (3.26)

with $\tilde{P}_i(n) := P_i(n) - n$ for all $i \in [k]$. By (3.2) and $\deg P_1 \ge 2$, $\deg \tilde{P}_i = \deg P_i$ for each $i \in [k]$. Denote

$$H(x) := e(rx/q)h(x), \qquad x \in \mathbb{Z}$$

Applying (3.26), e(rn/q) = e(r(x+n)/q)e(-rx/q) and Fubini's Theorem, we deduce

$$\|\mathbb{E}_{n \in J_N} f_1(x - \tilde{P}_1(n) \cdots f_k(x - \tilde{P}_k(n)) \ H(x + n)\|_{\ell^1(x \in \mathbb{Z})} \gtrsim \delta^{O(1)} N^{d_k}$$

which, combined with the dual arguments and the support condition for h, yields that there exists a 1-bounded function $h_0 \in \ell^{\infty}(\mathbb{Z})$ supported on $[\pm N_{**}]$ with $N_{**} \sim N^{d_k}$ such that

$$|\langle \mathcal{A}_N^{\tilde{\mathcal{P}}}(H, f_1, \dots, f_k), h_0 \rangle| \gtrsim \delta^{O(1)} N^{d_k},$$

where $\tilde{\mathcal{P}}(n) = (-n, \tilde{P}_1(n), \dots, \tilde{P}_k(n))$ is a map from $\mathbb{Z} \to \mathbb{Z}^{k+1}$ and $\mathcal{A}_N^{\tilde{\mathcal{P}}}$ is given by

$$\mathcal{A}_N^{\tilde{\mathcal{P}}}(H, f_1, \dots, f_k)(x) := \mathbb{E}_{n \in J_N} H(x+n) \ f_1(x-\tilde{P}_1(n)) \cdots f_k(x-\tilde{P}_k(n)).$$

Note that $\tilde{\mathcal{P}}$ is a family of polynomials with distinct degrees as well as the polynomial with the lowest degree in $\tilde{\mathcal{P}}$ is linear; and H, f_1, \ldots, f_k, h_0 are all 1-bounded functions supported on $[\pm N_{**}]$. As a result, we can obtain the conclusions in Theorem 3.3 for the case where deg $P_1 \geq 2$ by applying Theorem 3.2 with (k, N_*) replaced by $(k+1, N_{**})$.

We will now prove Proposition 3.4 by generalizing the method used in [32] (which addresses the special bilinear case) to the more general multilinear case. For convenience, we adopt

$$\Lambda_{\operatorname{Cr},\omega} := \Lambda_{\operatorname{Cram\'er},\omega}$$
 whenever $\omega \geq 1$.

Proof of Proposition 3.4. Without loss of generality, we assume that δ is sufficiently small. By applying Lemma 2.3 and (2.23) in in order, for any $z_1 \sim z_2$ with $1 \le z_1, z_2 \le \exp(\log^{1/10} N)$,

$$N^{-d_k} |\langle \tilde{A}_{N,\Lambda_{\mathrm{Cr},z_1} - \Lambda_{\mathrm{Cr},z_2}}(f_1, \cdots, f_k), h \rangle| \lesssim \langle \operatorname{Log} z_1 \rangle \left(N^{-1} + \|\Lambda_{\mathrm{Cr},z_1} - \Lambda_{\mathrm{Cr},z_2}\|_{u^{d_k+1}} \right)^{1/K}$$

$$\lesssim z_1^{-c_0}$$
(3.27)

for some $c_{\circ} = c_{\circ}(d_k) > 0$, where we have used $|\Lambda_{\operatorname{Cr},z_1} - \Lambda_{\operatorname{Cr},z_2}| \lesssim \langle \operatorname{Log} z_1 \rangle$ derived from (2.18). Then, for any $1 \leq \omega \leq \exp(\operatorname{Log}^{1/C_0} N)$, we can deduce from (3.27) that

$$|\langle \tilde{A}_{N,\Lambda_N-\Lambda_{\mathrm{Cr},\omega}}(f_1,\cdots,f_k),h\rangle| \leq \sum_{n=\mathrm{Log}\,\omega}^{\infty} |\langle \tilde{A}_{N,\Lambda_{\mathrm{Cr},2^{n+1}}-\Lambda_{\mathrm{Cr},2^n}}(f_1,\cdots,f_k),h\rangle| + O(\omega^{-c_\circ}N^{d_k})$$

$$\lesssim (\omega^{-c_\circ} + \sum_{n=\mathrm{Log}\,\omega}^{\infty} 2^{-c_\circ n})N^{d_k} \lesssim \omega^{-c_\circ}N^{d_k}.$$
(3.28)

Thus, to prove (3.24), it suffices to show that there exists $c_1 > 0$ such that

$$|\langle \tilde{A}_{N,\Lambda_{\mathrm{HB},\omega}-\Lambda_{\mathrm{Cr},\omega}}(f_1,\cdots,f_k),h\rangle| \lesssim \omega^{-c_1}N^{d_k}$$
 (3.29)

for all $1 \le \omega \le \exp(\operatorname{Log}^{1/C_0} N)$. Indeed, by setting

$$\omega = \omega_0 := \delta^{-2/\min\{c_0, c_1\}}$$

(since (3.22)), and applying the triangle inequality along with (3.29) and (3.28), we can finish the proof of (3.24) using this ω_{\circ} . To prove (3.29), we shall use Lemma 2.3 again. However, since this weight $\Lambda_{\mathrm{HB},\omega}$ lacks a suitable upper bound like $\Lambda_{\mathrm{Cr},\omega}$, we cannot apply Lemma 2.3 directly. To bridge this gap, we need to introduce a modified weight $\Lambda_{\mathrm{HB},\omega}^{\epsilon}$ defined by

$$\Lambda_{\mathrm{HB},\omega}^{\epsilon}(n) := \Lambda_{\mathrm{HB},\omega}(n) \ \mathbb{1}_{|\Lambda_{\mathrm{HB},\omega}(n)| \leq \omega^{c_{\circ}\epsilon}}$$

with sufficiently small ϵ fixed later, where $1 \le \omega \le \exp(\operatorname{Log}^{1/C_0} N)$ and c_{\circ} is given by (3.27). So

$$|\Lambda_{\mathrm{HB},\omega}^{\epsilon}(n)| \le \omega^{c_0 \epsilon},$$
 (3.30)

and $\Lambda_{\mathrm{HB},\omega}^{\epsilon}$ vanishes on $\{n \in \mathbb{Z} : |\Lambda_{\mathrm{HB},\omega}(n)| > \omega^{c_0 \epsilon}\}$, which, with (2.19) $(k = k_{\epsilon} := 2 + \lfloor \epsilon^{-1} \rfloor)$, gives

$$\langle \log \omega \rangle^{2^{k_{\epsilon}} + k_{\epsilon}} \gtrsim \mathbb{E}_{n \in [N]} |\Lambda_{\mathrm{HB}, \omega}(n)|^{k_{\epsilon}} \gtrsim \mathbb{E}_{n \in [N]} |D_{\mathrm{HB}, \omega}(n)|^{k_{\epsilon}} \mathbb{1}_{|\Lambda_{\mathrm{HB}, \omega}(n)| > \omega^{c_{\circ} \epsilon}}$$
$$\gtrsim \omega^{c_{\circ}} \mathbb{E}_{n \in [N]} |D_{\mathrm{HB}, \omega}(n)|,$$

where $D_{\mathrm{HB},\omega} := \Lambda_{\mathrm{HB},\omega} - \Lambda_{\mathrm{HB},\omega}^{\epsilon}$; consequently, we have

$$\mathbb{E}_{n \in [N]} |D_{\mathrm{HB},\omega}(n)| \lesssim \omega^{-c_0} \langle \mathrm{Log}\,\omega \rangle^{O_{\epsilon}(1)}. \tag{3.31}$$

Then, since (3.31) and $1 \le \omega \le \exp(\operatorname{Log}^{1/C_0} N)$, we immediately obtain a crude estimate

$$|\langle \tilde{A}_{N,D_{\mathrm{HB},\omega}}(f_1,\cdots,f_k),h\rangle| \lesssim \mathbb{E}_{n\in[N]}|D_{\mathrm{HB},\omega}(n)| \lesssim_{\epsilon} \omega^{-c_{\circ}}N.$$
(3.32)

By applying (3.31) and (2.24), there exists a positive constant $c'_{\circ} = c'_{\circ}(d_k)$ such that

$$\|\Lambda_{\mathrm{Cr},\omega} - \Lambda_{\mathrm{HB},\omega}^{\epsilon}\|_{u^{d_{k}+1}[N]} \leq \|\Lambda_{\mathrm{Cr},\omega} - \Lambda_{\mathrm{HB},\omega}\|_{u^{d_{k}+1}[N]} + \mathbb{E}_{n\in[N]}|D_{\mathrm{HB},\omega}(n)|$$

$$\lesssim \omega^{-c'_{\circ}} + \omega^{-c_{\circ}} \langle \operatorname{Log}\omega \rangle^{O_{\epsilon}(1)}.$$
(3.33)

In addition, by the upper bounds in (2.18) and (3.30), we have

$$|\Lambda_{\mathrm{Cr},\omega} - \Lambda_{\mathrm{HB},\omega}^{\epsilon}| \lesssim \omega^{c_0\epsilon} + \mathrm{Log}\,\omega \lesssim_{\epsilon} \omega^{c_0\epsilon}. \tag{3.34}$$

Combining Lemma 2.3 with (3.33)-(3.34), we obtain

$$N^{-d_{k}} |\langle \tilde{A}_{N,\Lambda_{\mathrm{Cr},\omega} - \Lambda_{\mathrm{HB},\omega}^{\epsilon}}(f_{1}, \cdots, f_{k}), h \rangle| \lesssim \omega^{c_{\circ}\epsilon} (N^{-1} + \|\Lambda_{\mathrm{Cr},\omega} - \Lambda_{\mathrm{HB},\omega}^{\epsilon}\|_{u^{d_{k}+1}[N]})^{1/K}$$

$$\lesssim \omega^{c_{\circ}\epsilon} (\omega^{-c'_{\circ}/K} + \langle \operatorname{Log} \omega \rangle^{O_{\epsilon}(1)} \omega^{-c_{\circ}/K}).$$
(3.35)

Set $\epsilon = \min\{1/(2K), c'_{\circ}/(2c_{\circ}K)\}$. By (3.35) and (3.32), we have

$$|\langle \tilde{A}_{N,\Lambda_{\mathrm{Cr},\omega}-\Lambda_{\mathrm{HB},\omega}}(f_1,\cdots,f_k),h\rangle| \lesssim \omega^{-\min\{c_\circ,c_\circ'\}/(4K)} N^{d_k}, \tag{3.36}$$

which yields (3.29) with $c_1 = \min\{c_\circ, c_\circ\}/(4K)$. This ends the proof of Proposition 3.4. \square

3.2. Structure of dual functions. With Theorem 3.3 in hand, we can obtain the structure of the dual function $\tilde{A}_{N,\Lambda_N}^{*j}(f_1,\ldots,f_k)$ for each $j\in[k]$, as defined in (3.6).

Theorem 3.5. Let N, k, \mathcal{P}, N_* and δ be given as in Theorem 3.3. Fix $j \in [k]$. Then for all 1-bounded functions $f_1, \ldots, f_k \in S(\mathbb{Z})$ supported on $[\pm N_*]$, we have

$$\tilde{A}_{N,\Lambda_N}^{*j}(f_1,\ldots,f_k) = \sum_{\alpha_j \in \mathbb{Q}/\mathbb{Z}: \ h(\alpha_j) \lesssim \delta^{-O(1)}} F_{\alpha_j} + E_1 + E_2, \tag{3.37}$$

where each $F_{\alpha_j} \in \ell^2(\mathbb{Z})$ has Fourier transform supported in the $O(\delta^{-O(1)}/N^{d_j})$ -neighborhood of $\alpha_j \in \mathbb{Q}/\mathbb{Z}$ and satisfies the estimates

$$||F_{\alpha_i}||_{\ell^{\infty}(\mathbb{Z})} \lesssim \delta^{-O(1)} \quad and \quad ||F_{\alpha_i}||_{\ell^1(\mathbb{Z})} \lesssim \delta^{-O(1)} N^{d_k},$$
 (3.38)

and the error terms $E_1 \in \ell^1(\mathbb{Z})$ and $E_2 \in \ell^2(\mathbb{Z})$ satisfy the following estimates:

$$||E_1||_{\ell^1(\mathbb{Z})} \le \delta N^{d_k} \quad and \quad ||E_2||_{\ell^2(\mathbb{Z})} \le \delta.$$
 (3.39)

Proof of Theorem 3.5. Combining Theorem 3.3 with the Hahn-Banach theorem (see [31, Lemma 6.9] for the details), we can obtain desired results by following the arguments yielding [31, Corollary 6.10] line by line.

By using the decomposition (3.37) for each $j \in [k]$ and following the approach in [31], we obtain two useful ℓ^2 bounds for $\tilde{A}_{N,\Lambda_N}^{*j}$ in the following two propositions. As a matter of fact, we will extend the approach for the bilinear case in [31] and employ Theorem 2.1.

Proposition 3.6. Fix
$$k \in \mathbb{Z}_+$$
 and $j \in [k]$. Let $N, N_* \ge 1$ with $N_* \sim N^{d_k}$, and $l \in \mathbb{N}$ with $c' \operatorname{Log}^{1/C_0} N \ge l$ (3.40)

where c' is given by (3.22). Then there exists a positive c_0 depending only on c' such that

$$\|(1 - \Pi_{\leq l, \leq d_j(N, l)}) \tilde{A}_{N, \Lambda_N}^{*j}(f_1, \dots, f_k)\|_{\ell^2(\mathbb{Z})} \lesssim 2^{-c_0 l} N^{d_k/2} \|f_1\|_{\ell^\infty(\mathbb{Z})} \dots \|f_k\|_{\ell^\infty(\mathbb{Z})}$$
(3.41)

with $d_j(N,l) := -d_j \operatorname{Log} N + d_j l$, whenever $f_1, \ldots, f_k \in \mathbf{S}(\mathbb{Z})$ are supported on $[-N_*, N_*]$.

Proof of Proposition 3.6. We can assume N, l large enough since the desired result can be obtained by (2.21) and Hölder's inequality otherwise. By normalization, we set $||f_l||_{\ell^{\infty}(\mathbb{Z})} = 1$ for all $l \in [k]$. Thus, to achieve the goal (3.41), it suffices to show

$$\|(1 - \prod_{\leq l, \leq d_j(N,l)}) \tilde{A}_{N,\Lambda_N}^{*j}(f_1, \dots, f_k)\|_{\ell^2(\mathbb{Z})} \lesssim 2^{-c_0 l} N^{d_k/2}.$$
(3.42)

By Theorem 3.5 for $\delta = 2^{-c'_0 l}$ with c'_0 small enough such that (3.22) holds (using (3.40)), and

$$(1 - \prod_{\leq l, \leq d_i(N,l)}) F_{\alpha_j} = 0$$

for all F_{α_i} in the decomposition (3.37), we have

$$(1 - \prod_{\leq l, \leq d_j(N,l)}) \tilde{A}_{N,\Lambda_N}^{*j}(f_1, \dots, f_k) = (1 - \prod_{\leq l, \leq d_j(N,l)}) E_1 + (1 - \prod_{\leq l, \leq d_j(N,l)}) E_2.$$
(3.43)

By (3.39), we have $\|(1 - \prod_{\leq l, \leq d_j(N,l)})E_2\|_{\ell^2(\mathbb{Z})} \leq \|E_2\|_{\ell^2(\mathbb{Z})} \lesssim \delta$. Hence, to prove (3.42), it suffices to show that for some $c_0 > 0$,

$$\|(1 - \prod_{\leq l, \leq d_j(N,l)}) E_1\|_{\ell^2(\mathbb{Z})} \lesssim 2^{-c_0 l} N^{d_k/2}. \tag{3.44}$$

Since l is sufficiently large, by interpolation, we further reduce the proof of (3.44) to proving

$$\|(1 - \prod_{\leq l, \leq d_j(N,l)}) E_1\|_{\ell^q(\mathbb{Z})} \lesssim 2^{\mathbf{C}_{2\lfloor q'\rfloor}(2^l)} \delta^{1/2} N^{d_k/q}$$
 and (3.45)

$$\|(1 - \prod_{\leq l, \leq d_j(N, l)}) E_1\|_{\ell^{q'}(\mathbb{Z})} \lesssim 2^{\mathbf{C}_{2\lfloor q'\rfloor}(2^l)} N^{d_k/q'} \qquad (q' = q/(q-1))$$
(3.46)

for some $q \in (1,2)$, where $\mathbf{C}_{2\lfloor q' \rfloor}(2^l)$ is given by (2.15). We will prove (3.45) and (3.46) in order.

By triangle inequality, (3.37), (2.21), (3.38) and (3.39), we deduce $||E_1||_{\ell^{\infty}(\mathbb{Z})} \lesssim \delta^{-O(1)}$. Interpolating this inequality with the estimate of E_1 in (3.39) gives $||E_1||_{\ell^{\tilde{q}}(\mathbb{Z})} \lesssim \delta^{1/2} N^{d_k/\tilde{q}}$ for some $\tilde{q} \in (1,2)$ close to 1. By using (2.14), we then obtain that for any $\epsilon > 0$,

$$\|(1 - \Pi_{\leq l, \leq d_j(N, l)}) E_1\|_{\ell^{\tilde{q}}(\mathbb{Z})} \lesssim 2^{\mathbf{C}_{2 \lfloor \tilde{q}' \rfloor}(2^l)} \|E_1\|_{\ell^{\tilde{q}}(\mathbb{Z})} \lesssim 2^{\mathbf{C}_{2 \lfloor \tilde{q}' \rfloor}(2^l)} \delta^{1/2} N^{d_k/\tilde{q}}, \tag{3.47}$$

which gives (3.45) with $q = \tilde{q}$. Next, we prove (3.46) with $q = \tilde{q}$. Noting that $\tilde{A}_{N,\Lambda_N}^{*j}(f_1,\ldots,f_k)$ is bounded by O(1) (since (2.21)) and is supported on $[-\tilde{N}_*,\tilde{N}_*]$ with $\tilde{N}_* \sim N^{d_k}$, we have a direct estimate $\|\tilde{A}_{N,\Lambda_N}^{*j}(f_1,\ldots,f_k)\|_{\ell^{\tilde{q}'}(\mathbb{Z})} \lesssim N^{d_k/\tilde{q}'}$, which, by (2.14), implies

$$\|(1-\Pi_{\leq l,\leq d_i(N,l)})\tilde{A}_{N,\Lambda_N}^{*j}(f_1,\ldots,f_k)\|_{\ell^{\tilde{q}'}(\mathbb{Z})} \lesssim 2^{\mathbf{C}_{2\lfloor \tilde{q}'\rfloor}(2^l)}N^{d_k/\tilde{q}'}.$$

This, along with the direct inequality $\|(1 - \prod_{\leq l, \leq d_j(N,l)})E_2\|_{\ell^{\tilde{q}'}(\mathbb{Z})} \leq \|E_2\|_{\ell^2(\mathbb{Z})} \lesssim \delta$ as well as (3.43), gives the desired (3.46) with $q = \tilde{q}$. We end the proof of Proposition 3.6.

Proposition 3.7. Let k, j, N, N_*, l and the notation $d_j(N, l)$ be given as in Proposition 3.6. Let $\bar{\iota} : [k] \to \{1, k\}$ be a map obeying $\bar{\iota}(m) = k$ if $m \in [k-1]$ and $\bar{\iota}(k) = 1$. Then there exists some $c_0 > 0$ such that

$$\|(1 - \prod_{\leq l, \leq d_j(N, l)}) \tilde{A}_{N, \Lambda_N}^{*j}(f_1, \dots, f_k)\|_{\ell^2(\mathbb{Z})} \lesssim 2^{-c_0 l} \|f_{\bar{\iota}(j)}\|_{\ell^2(\mathbb{Z})} \prod_{i \in [k] \setminus \{\bar{\iota}(j)\}} \|f_i\|_{\ell^\infty(\mathbb{Z})}$$
(3.48)

whenever $f_{\bar{\iota}(j)} \in \ell^2(\mathbb{Z})$ and $\mathbb{1}_{i \in [k] \setminus \bar{\iota}(j)}$ $f_i \in \ell^\infty(\mathbb{Z})$.

Remark 3. If $k \geq 3$, the definition $\bar{\iota}(k) = 1$ can be replaced by $\bar{\iota}(k) = m$ with $m \in [k] \setminus \{1, k\}$.

The proof of Proposition 3.7 requires the weighted L^p improving bounds below.

Lemma 3.8. Let $N \ge 1$, let Q be a polynomial with integer coefficients obeying deg $Q = d \ge 1$. Then there exists some $q_0 \in (1,2)$ such that

$$\|\mathbb{E}_{n\in[N]}(\Lambda(n) + \Lambda_N(n))|H(\cdot - Q(n))|\|_{\ell^2(\mathbb{Z})} \lesssim N^{d(1/2 - 1/q_0)} \|H\|_{\ell^{q_0}(\mathbb{Z})} \quad \text{and}$$

$$\|\mathbb{E}_{n\in[N]}(\Lambda(n) + \Lambda_N(n))|H(\cdot - Q(n))|\|_{\ell^{q'_0}(\mathbb{Z})} \lesssim N^{d(1/q'_0 - 1/q_0)} \|H\|_{\ell^{q_0}(\mathbb{Z})},$$
(3.49)

where $q'_0 = q_0/(q_0 - 1)$.

We refer to [32, Lemma 5.1] for the proof of (3.49) depending on the estimate (2.23).

Proof of Proposition 3.7. Since the linear case k=1 can be addressed using a standard procedure (Plancherel's identity and the exponent sum estimate), we will concentrate on the case where $k \geq 2$ in the following discussion. We begin by proving (3.48) under the assumption that the functions $f_1, \ldots, f_k \in S(\mathbb{Z})$ are supported on $[-N_*, N_*]$ with $N_* \sim N^{d_k}$. By (3.49)₁

with $(H,d) = (f_{\bar{\iota}(j)}, d_k)$ and $Q = (P_k - P_j) \mathbb{1}_{j \in [k-1]} + (P_1 - P_k) \mathbb{1}_{j=k}$, there exists $q_0 \in (1,2)$ such that

$$\|\tilde{A}_{N,\Lambda_N}^{*j}(f_1,\ldots,f_k)\|_{\ell^2(\mathbb{Z})} \lesssim N^{d_k(1/2-1/q_0)} \|f_{\bar{\iota}(j)}\|_{\ell^{q_0}(\mathbb{Z})} \prod_{i\in[k]\setminus\{\bar{\iota}(j)\}} \|f_i\|_{\ell^{\infty}(\mathbb{Z})},$$

which, with (3.41), gives the desired (3.48) (under the support assumptions) by interpolation. Next, we relax the support assumptions of f_1, \ldots, f_k . We may assume that l is large enough since the goal is trivial otherwise. Following the arguments yielding an off diagonal bound (see

since the goal is trivial otherwise. Following the arguments yielding an off-diagonal bound (see [31, (5.19)]), we can derive a similar off-diagonal bound, which can be stated as follows: if f is supported on an interval I with $|I| \sim N^{d_k}$, then

$$\|\Pi_{\leq l, \leq d_j(N,l)} f\|_{\ell^2(J)} \lesssim 2^{\mathbf{C}_2(2^l)} \langle 2^{d_j(N,l)} \operatorname{dist}(I,J) \rangle^{-10} \|f\|_{\ell^2(I)},$$
 (3.50)

where the interval J satisfies $|J| \sim N^{d_k}$, and $2^{\mathbf{C}_2(2^l)}$ is given by (2.15). With the off-diagonal bound (3.50) in hand, we can eliminate the support conditions of f_1, \ldots, f_k by the following standard process. Decompose $f_{\bar{\iota}(j)} = \sum_{I \in \mathcal{I}} f_{\bar{\iota}(j)} \mathbb{1}_I$, where I ranges over a partition \mathcal{I} of \mathbb{R} into intervals I of length N^{d_k} . Note that

$$\mathcal{D}_{I}(f_{\bar{\iota}(j)}) := \tilde{A}_{N,\Lambda_{N}}^{*,j}(f_{1},\ldots,f_{j-1},f_{\bar{\iota}(j)}\mathbb{1}_{I},f_{j+1},\ldots,f_{k})$$

is supported in an $O(N^{d_k})$ -neighborhood of I. We normalize $||f_i||_{\ell^{\infty}(\mathbb{Z})} = 1$ for all $i \in [k] \setminus \{\bar{\iota}(j)\}$. Then, by the previous discussion (that is, (3.48) with the support condition holds), there exists c > 0 such that for any $I \in \mathcal{I}$,

$$\|(1 - \Pi_{\leq l, \leq d_j(N,l)}) \mathcal{D}_I(f_{\bar{\iota}(j)})\|_{\ell^2(\mathbb{Z})} \lesssim 2^{-cl} \|f_{\bar{\iota}(j)}\|_{\ell^2(I)}$$
(3.51)

with the implicit constant independent of l. To obtain the desired result, it suffices to prove

$$\| \sum_{I \in \mathcal{I}} (1 - \prod_{\leq l, \leq d_j(N, l)}) \mathcal{D}_I(f_{\bar{\iota}(j)}) \|_{\ell^2(\mathbb{Z})} \lesssim 2^{-c_0 l} \|f_{\bar{\iota}(j)}\|_{\ell^2(\mathbb{Z})}$$
(3.52)

for some $c_0 > 0$. By writing the left-hand side of (3.52) as the inner product form, we can achieve (3.52) by combining (3.51) and the off-diagonal bound (3.50).

3.3. **Proof of Theorem 3.1.** As stated in *Remark* 2 below Theorem 3.1, we can assume that $l \leq \text{Log}^{1/C_0} N$ and l, N are sufficiently large. Then it suffices to show

$$\|\tilde{A}_{N,\Lambda_N}^{\mathcal{P}}(f_1,\ldots,f_k)\|_{\ell^q(\mathbb{Z})} \lesssim 2^{-cl} \|f_1\|_{\ell^{q_1}(\mathbb{Z})} \cdots \|f_k\|_{\ell^{q_k}(\mathbb{Z})}$$
 (3.53)

Note that (2.21) and Hölder's inequality give (3.53) without the factor 2^{-cl} . Then, by interpolation, it is sufficient for (3.53) to show that there exists $\tilde{c} > 0$ such that

$$\|\tilde{A}_{N,\Lambda_N}^{\mathcal{P}}(f_1,\ldots,f_k)\|_{\ell^1(\mathbb{Z})} \lesssim 2^{-\tilde{c}l} \|f_j\|_{\ell^2(\mathbb{Z})} \|f_{\bar{\iota}(j)}\|_{\ell^2(\mathbb{Z})} \prod_{i\in[k]\setminus\{j,\bar{\iota}(j)\}} \|f_i\|_{\ell^\infty(\mathbb{Z})}$$
(3.54)

with the map $\bar{\iota}$ defined as in Proposition 3.7. By applying dual arguments and (3.5), to prove (3.54), it suffices to show that for all function h with $||h||_{\ell^{\infty}(\mathbb{Z})} = 1$,

$$\left| \left\langle \tilde{A}_{N,\Lambda_N}^{\mathcal{P},*j}(f_1,\ldots,f_{j-1},h,f_{j+1},\ldots,f_k),f_j \right\rangle \right|$$

$$\lesssim 2^{-\tilde{c}l} \|f_j\|_{\ell^2(\mathbb{Z})} \|f_{\bar{\iota}(j)}\|_{\ell^2(\mathbb{Z})} \prod_{i \in [k] \setminus \{j,\bar{\iota}(j)\}} \|f_i\|_{\ell^{\infty}(\mathbb{Z})}.$$

$$(3.55)$$

Using the support condition for f_i , we rewrite the inner product on the left-hand side as

$$\langle (1 - \prod_{\leq l, \leq d_j(N,l)}) \tilde{A}_{N,\Lambda_N}^{\mathcal{P},*j}(f_1, \dots, f_{j-1}, h, f_{j+1}, \dots, f_k), f_j \rangle,$$
 (3.56)

with $d_j(N, l)$ given as in Proposition 3.6. Finally, by applying Hölder's inequality and (3.48) to the above inner product (3.56), we can achieve (3.55).

4. Proof of Theorem 1.3: reducting to major arcs

In this section, we will reduce the proof of Theorem 1.3 to establishing the major arcs estimates, and subsequently provide the reduction for these estimates.

4.1. Reduction of Theorem 1.3 by multilinear Weyl inequality. We shall use the minor arcs estimates established in Theorem 3.1 to reduce the proof of (1.9) to showing the following theorem. Let $\{I_n(d)\}_{n\in\mathbb{N}}$ denote the sets given by⁶

$$I_n(d) := [2\rho_d^n, 2\rho_d^{n+1}), \qquad d \in \mathbb{Z}_+,$$
(4.1)

where $\rho_d = d/(d-1)$ if $d \ge 2$, and $\rho_1 = 2$.

Theorem 4.1. Let $k \geq 2$, \mathcal{P} be a polynomial mapping satisfying (3.1) and (3.2), $n \in \mathbb{N}$ and $q \in I_n(d_k)$. Let $\{q_{jv}\}_{(j,v)\in[k]\times\{1,2\}} \subset (1,\infty)$ satisfying $\frac{1}{q_{j1}} + \sum_{i\in[k]\setminus\{j\}} \frac{1}{q_{i2}} = \frac{1}{q}$ for $j \in [k]$ and $q_* \in I_n(d_k)$, where $q_* = \max\{q_{j1}, j \in [k]\}$. Then, for any r > 2, we have

$$\| \left(\tilde{A}_{N,\Lambda_{N}}^{\mathcal{P}} (\Pi_{\leq l_{(N)}, \leq -d_{1}(\operatorname{Log} N - l_{(N)})} f_{1}, \cdots, \Pi_{\leq l_{(N)}, \leq -d_{k}(\operatorname{Log} N - l_{(N)})} f_{k}) \right)_{N \in \mathbb{D}} \|_{\ell^{q}(\mathbb{Z}; \mathbf{V}^{r})}$$

$$\lesssim \sum_{j \in [k]} \left(\| f_{j} \|_{L^{q_{j_{1}}}(\mathbb{Z})} \prod_{i \in [k] \setminus \{j\}} \| f_{i} \|_{L^{q_{i_{2}}}(\mathbb{Z})} \right)$$

$$(4.2)$$

for any $f_j \in \bigcap_{v=1,2} \ell^{q_{jv}}(\mathbb{Z})$ with $j \in [k]$, where the scale $l_{(N)}$ is defined by

$$l_{(N)} := \lfloor \tilde{c} \operatorname{Log}^{1/C_0} N \rfloor \tag{4.3}$$

with sufficiently small $\tilde{c} > 0$.

Remark 4. We will reduce the proof of Theorem 4.1 to establishing Propositions 4.4-4.6. In particular, the restrictions for the parameters q, $\{q_{jv}\}_{(j,v)\in[k]\times\{1,2\}}$ and the functions f_1,\ldots,f_k in Theorem 4.1 come from the conditions in Proposition 4.6; see Remark 6 in Subsection 4.3.2 for more details.

Set $p_0 \in 2\mathbb{Z}_+$ large enough such that the parameters $q, \tilde{c}, q_{11}, q_{12}, \dots, q_{k1}, q_{k2}$ in Theorem 4.1 satisfies

$$q, \tilde{c}, q_{11}, q_{12}, \dots, q_{k1}, q_{k2} \in [p'_0, p_0].$$
 (4.4)

By Minkowski's inequality, (2.6), (2.18) and Hölder's inequality, it is easy to obtain (1.9) with the lacunary set \mathbb{D} replaced by its subset $\mathbb{D} \cap \{N \leq 1\}$. Then, by (2.5), it suffices to prove (1.9) with the set \mathbb{D} replaced by

$$\mathbb{D}^B := \mathbb{D} \cap \{ N \ge \mathcal{C}_0 \}, \tag{4.5}$$

where C_0 (depending on p_0 and C_0 given as in Section 3) is sufficiently large. In fact, the above $l_{(N)}$ is selected according to the decay factor on the right-hand side of (3.4). Below we prove (1.9) under the assumption that Theorem 4.1 holds.

Proof of Theorem 1.3 (accepting Theorem 4.1). Since \mathbb{D} is lacunary, by the triangle inequality and (2.6), it suffices to show the following two inequalities: first,

$$\|\tilde{A}_{N,\Lambda-\Lambda_N}^{\mathcal{P}}(f_1,\dots,f_k)\|_{\ell^q(\mathbb{Z})} \lesssim \langle \operatorname{Log} N \rangle^{-2} \prod_{i \in [k]} \|f_i\|_{\ell^{q_i}(\mathbb{Z})}$$
(4.6)

for any $1/q_1 + \cdots + 1/q_k = 1/q \le 1$ with $q_1, \dots, q_k \in (1, \infty)$; and second,

$$\| \left(\tilde{A}_{N,\Lambda_N}^{\mathcal{P}}(f_1, \cdots, f_k) \right)_{N \in \mathbb{D}^B} \|_{\ell^q(\mathbb{Z}; \mathbf{V}^r)} \lesssim \sum_{j \in [k]} \left(\| f_j \|_{L^{q_{j1}}(\mathbb{Z})} \prod_{i \in [k] \setminus \{j\}} \| f_i \|_{L^{q_{i2}}(\mathbb{Z})} \right)$$
(4.7)

⁶The choices for $I_n(d)$ and ρ_d are both non-unique.

with q, $\{q_{jv}\}_{(j,v)\in[k]\times\{1,2\}}$, r given as in Theorem 4.1. We will prove (4.6) and (4.7) sequentially. In particular, the proof of (4.6) extends the methodology developed in [32], adapting it to our broader framework.

By (2.18), (2.21), (3.3) and Hölder's inequality, we have

$$\|\tilde{A}_{N,\Lambda-\Lambda_N}^{\mathcal{P}}(f_1,\cdots,f_k)\|_{\ell^q(\mathbb{Z})} \leq \sum_{\omega\in\{\Lambda,\Lambda_N\}} \|\tilde{A}_{N,\omega}^{\mathcal{P}}(|f_1|,\cdots,|f_k|)\|_{\ell^q(\mathbb{Z})} \lesssim \prod_{i\in[k]} \|f_i\|_{\ell^{q_i}(\mathbb{Z})},$$

which implies, by interpolation, that it is sufficient for (4.6) to show

$$\|\tilde{A}_{N,\Lambda-\Lambda_N}^{\mathcal{P}}(f_1,\dots,f_k)\|_{\ell^1(\mathbb{Z})} \lesssim_M \langle \text{Log } N \rangle^{-M} \|f_1\|_{\ell^2(\mathbb{Z})} \|f_k\|_{\ell^2(\mathbb{Z})} \prod_{i \in [k] \setminus \{1,k\}} \|f_i\|_{\ell^\infty(\mathbb{Z})}$$
(4.8)

for any sufficiently large $M \geq 1$. To prove (4.8), we observe that by dual arguments, it suffices to show that for any $||h||_{\ell^{\infty}(\mathbb{Z})} = ||f_i||_{\ell^{\infty}(\mathbb{Z})} = 1$ with $i \in [k] \setminus \{1, k\}$,

$$\|\tilde{A}_{N,\Lambda-\Lambda_N}^{\mathcal{P},*1}(h,f_2,\cdots,f_k)\|_{\ell^2(\mathbb{Z})} \lesssim_M \langle \operatorname{Log} N \rangle^{-M} \|f_k\|_{\ell^2(\mathbb{Z})}. \tag{4.9}$$

Moreover, by localization, it suffices to prove (4.9) for the functions h, f_2, \dots, f_k supported in $[-N_*, N_*]$ with $N_* \sim N^{d_k}$. From [35, Theorem 1.1]⁷ we deduce that for any large $M \geq 1$,

$$\|\Lambda - \Lambda_N\|_{u^{d_k + 1}[N]} \lesssim_{M, C_0} \langle \operatorname{Log} N \rangle^{-M}, \tag{4.10}$$

where $\|\cdot\|_{u^{d_k+1}[N]}$ is defined by (2.20) with $s=d_k$. Using dual arguments, Lemma 2.3 and (4.10), we infer that for any $M \geq 1$,

$$N^{-d_{k}} \|\tilde{A}_{N,\Lambda-\Lambda_{N}}^{\mathcal{P},*1}(h,f_{2},\cdots,f_{k})\|_{\ell^{1}(\mathbb{Z})} \leq (N^{-1} + \|\Lambda-\Lambda_{N}\|_{u^{d_{k}+1}[N]})^{1/K} \|f_{k}\|_{\ell^{\infty}(\mathbb{Z})}$$

$$\lesssim_{M} \langle \operatorname{Log} N \rangle^{-M} \|f_{k}\|_{\ell^{\infty}(\mathbb{Z})}.$$
(4.11)

By $(3.49)_2$ with $(H, d, Q) = (f_k, d_k, P_k - P_1)$, there exists some constant $\tilde{q} \in (1, 2)$ such that

$$\|\tilde{A}_{N,\Lambda-\Lambda_N}^{\mathcal{P},*1}(h,f_2,\cdots,f_k)\|_{\ell^{\tilde{q}'}(\mathbb{Z})} \leq \|\mathbb{E}_{n\in[N]}(\Lambda(n)+\Lambda_N(n))|f_k(\cdot+P_1(n)-P_k(n))|\|_{\ell^{\tilde{q}'}(\mathbb{Z})}$$

$$\lesssim N^{d_k(1/\tilde{q}'-1/\tilde{q})}\|f_k\|_{\ell^{\tilde{q}}(\mathbb{Z})}, \quad \text{where } \tilde{q}'=\tilde{q}/(\tilde{q}-1).$$

$$(4.12)$$

Interpolating (4.11) with (4.12), we obtain (4.9) by setting M large enough. This completes the proof of (4.6).

Next, we prove (4.7). For convenience, we denote

$$\Pi^{1}_{\leq m, \leq n} := \Pi_{\leq m, \leq n} \quad \text{and} \quad \Pi^{2}_{\leq m, \leq n} := 1 - \Pi_{\leq m, \leq n}$$
(4.13)

for any $(m,n) \in \mathbb{Z}^2$. Given the notation introduced in (4.13), the proof of (4.7) reduces to proving the inequality

$$\| \left(\tilde{A}_{N,\Lambda_{N}}^{\mathcal{P}} (\Pi_{\leq l_{(N)},\leq -d_{1}(\log N - l_{(N)})}^{\kappa_{1}} f_{1}, \cdots, \Pi_{\leq l_{(N)},\leq -d_{k}(\log N - l_{(N)})}^{\kappa_{k}} f_{k}) \right)_{N \in \mathbb{D}^{B}} \|_{\ell^{q}(\mathbb{Z}; \mathbf{V}^{r})}$$

$$\lesssim \sum_{j \in [k]} \left(\| f_{j} \|_{L^{q_{j_{1}}}(\mathbb{Z})} \prod_{i \in [k] \setminus \{j\}} \| f_{i} \|_{L^{q_{i_{2}}}(\mathbb{Z})} \right)$$

$$(4.14)$$

for any $(\kappa_1, \ldots, \kappa_k) \in \{1, 2\}^k$, where $q, \{q_{jv}\}_{(j,v) \in [k] \times \{1, 2\}}, r$ are given as in Theorem 4.1. By (4.2), we can achieve (4.14) for the case where $\kappa_j = 1$ for all $j \in [k]$. For the remaining cases, we shall apply Theorem 3.1. Since \mathbb{D} is lacunary, it suffices to demonstrate that for every $(\kappa_1, \ldots, \kappa_k) \neq (1, \ldots, 1)$,

$$\|\tilde{A}_{N,\Lambda_N}^{\mathcal{P}}(\Pi_{\leq l_{(N)},\leq -d_1(\log N - l_{(N)})}^{\kappa_1} f_1, \cdots, \Pi_{\leq l_{(N)},\leq -d_k(\log N - l_{(N)})}^{\kappa_k} f_k)\|_{\ell^q(\mathbb{Z})} \lesssim \langle \log N \rangle^{-2}$$
 (4.15)

for any $1/q_1 + \cdots + 1/q_k = 1/q \le 1$ with $q_1, \dots, q_k \in [p'_0, p_0]$ and for any $||f_i||_{\ell^{q_i}(\mathbb{Z})} = 1$. Without loss of generality, we assume $\kappa_{j_0} = 2$ for some $j_0 \in [k]$. From the definitions (4.13)

⁷While [35] only treats the case $C_0 = 10$, their methods extend to arbitrarily large C_0 .

and (2.12), we observe that the Fourier transform of $\Pi^2_{\leq l_{(N)}, \leq -d_{j_0}(\text{Log }N-l_{(N)})}f_{j_0}$ vanishes on $\mathfrak{M}_{\leq l_{(N)}, \leq -d_{j_0}(\text{Log }N-l_{(N)})-1}$. Thus, since $N \in \mathbb{D}^B$, by applying Theorem 3.1 (with $j=j_0$), Theorem 2.1 and (4.3), we can deduce that the left-hand side of (4.15) is

$$\lesssim 2^{-cl_{(N)}} 2^{k \mathbf{C}_{p_0}(2^{l_{(N)}})} \lesssim 2^{-\frac{c}{2}l_{(N)}} \lesssim \langle \operatorname{Log} N \rangle^{-2},$$

where $\mathbf{C}_{p_0}(\cdot)$ is given by (2.15). This yields (4.15), thus concluding the proof of (4.7).

4.2. Reduction of Theorem 4.1 via dyadic decompositions. It suffices to prove (4.2) with the set \mathbb{D} replaced by \mathbb{D}^B . In this section, following the process in [31, p.1047], we will employ both an "arithmetic" dyadic decomposition and a "continuous" dyadic decomposition.

Invoke the definition (2.12), and define

$$\Pi_{m,\leq n} = \Pi_{\leq m,\leq n} - \Pi_{\leq m-1,\leq n}$$

with the convention $\Pi_{\leq -1,\leq n}=0$. Through the arithmetic decomposition

$$\Pi_{\leq m, \leq n} = \sum_{0 \leq m' \leq m} \Pi_{m', \leq n},$$

the averages in (4.2) become

$$\sum_{l_1 \in \mathbb{N}_{\leq l_{(N)}}} \cdots \sum_{l_k \in \mathbb{N}_{\leq l_{(N)}}} \tilde{A}_{N,\Lambda_N}^{\mathcal{P}} \left(\Pi_{l_1, \leq -d_1(\operatorname{Log} N - l_{(N)})} f_1, \dots, \Pi_{l_k, \leq -d_k(\operatorname{Log} N - l_{(N)})} f_k \right). \tag{4.16}$$

For each $i \in [k]$, we decompose the bump function $\eta_{\leq -d_i(\log N - l_N)}$ (defined in (2.1)) into non-oscillatory and highly oscillatory components as follows:

$$\eta_{\leq -d_i(\log N - l_{(N)})} = \eta_N^0 + \sum_{s_i \in [l_{(N)}]} \eta_N^{s_i}$$

where

$$\eta_N^{s_i} := \begin{cases} \eta_{\leq -d_i(\operatorname{Log} N - s_i)} - \eta_{\leq -d_i(\operatorname{Log} N - s_i + 1)} & \text{if } s_i \in \mathbb{Z}_+, \\ \eta_{\leq -d_i \operatorname{Log} N} & \text{if } s_i = 0. \end{cases}$$
(4.17)

Recall (2.13) and introduce the following notation

$$\mathfrak{Q}_{\leq n} := \mathcal{R}_{\leq 2^n} \quad \text{and} \quad \mathfrak{Q}_n := \mathcal{R}_{\leq 2^n} \setminus \mathcal{R}_{<2^{n-1}}, \quad n \in \mathbb{N}, \tag{4.18}$$

where we adopt the convention $\mathcal{R}_{<2^{-1}} = \emptyset$. Then, we denote

$$\tilde{\Pi}_{l_i,s_i}^N := T_{\mathbb{Z}}^{\mathfrak{Q}_{l_i}}[\eta_N^{s_i}]. \tag{4.19}$$

Combining the support property (4.17) and the definition (4.19), we reformulate the arithmetic decomposition (4.16) as:

$$\sum_{l_1,s_1\in\mathbb{N}_{\leq l_{(N)}}}\cdots\sum_{l_k,s_k\in\mathbb{N}_{\leq l_{(N)}}}\tilde{A}_{N,\Lambda_N}^{\mathcal{P}}\left(\tilde{\Pi}_{l_1,s_1}^Nf_1,\ldots,\tilde{\Pi}_{l_k,s_k}^Nf_k\right).$$

As a result, it is sufficient for Theorem 4.1 to prove that there exists $c_0 > 0$ such that

$$\| \left(\tilde{A}_{N,\Lambda_{N}}^{\mathcal{P}}(\tilde{\Pi}_{l_{1},s_{1}}^{N}f_{1},\cdots,\tilde{\Pi}_{l_{k},s_{k}}^{N}f_{k}) \right)_{N \in \mathbb{D}_{l,s}} \|_{\ell^{q}(\mathbb{Z};\mathbf{V}^{r})}$$

$$\lesssim 2^{-c_{0}(l+s)} \sum_{j \in [k]} \left(\|f_{j}\|_{L^{q_{j_{1}}}(\mathbb{Z})} \prod_{i \in [k] \setminus \{j\}} \|f_{i}\|_{L^{q_{i_{2}}}(\mathbb{Z})} \right)$$

$$(4.20)$$

with $q, \{q_{jv}\}_{(j,v)\in[k]\times\{1,2\}}, r$ given as in Theorem 4.1, where the set $\mathbb{D}_{l,s}$ is given by

$$\mathbb{D}_{l,s} = \{ N \in \mathbb{D}^B : \ l_{(N)} \ge \max\{l, s\} \} \qquad \text{with}$$

$$l = \max\{l_1, \dots, l_k\} \qquad \text{and} \qquad s = \max\{s_1, \dots, s_k\}.$$

$$(4.21)$$

4.3. Major arcs approximation and reduction of (4.20). We begin by establishing the necessary preliminaries for the major arcs approximation. For any $N \ge 1$, define the exponential sum and its continuous part by

$$m_{N,\Lambda_N}(\xi) := \frac{1}{\lfloor N \rfloor} \sum_{n \in J_N} \Lambda_N(n) \ e(\xi \cdot \mathcal{P}(n)) \qquad (\xi \in \mathbb{T}^k) \quad \text{and}$$

$$\tilde{m}_{N,\mathbb{R}}(\zeta) := \int_{1/2}^1 e(\zeta \cdot \mathcal{P}(Nt)) dt \qquad (\zeta \in \mathbb{R}^k),$$

$$(4.22)$$

where $J_N = [N] \setminus [N/2]$. Let $\mathfrak{Q}_{l_1, \dots, l_k}$ denote the set of rational fractions

$$\mathfrak{Q}_{l_1,\dots,l_k} := \{ a/q \in (\mathbb{T} \cap \mathbb{Q})^k : (a,q) = 1, \quad a/q \in \mathfrak{Q}_{l_1} \times \dots \times \mathfrak{Q}_{l_k} \}$$

$$(4.23)$$

where each \mathfrak{Q}_{l_i} is defined as in (4.18). For $a/q \in \mathfrak{Q}_{l_1,\cdots,l_k}$, define the exponential sum

$$G^{\times}(\frac{a}{q}) = \mathbb{E}_{n \in [q] \times} e(\frac{a}{q} \cdot \mathcal{P}(n)) = \frac{1}{\varphi(q)} \sum_{n \in [q] \times} e(\frac{a}{q} \cdot \mathcal{P}(n)). \tag{4.24}$$

Since $l = \max\{l_1, \dots, l_k\}$, we can see

$$2^l \le q \le 2^{kl} \tag{4.25}$$

whenever $a/q \in \mathfrak{Q}_{l_1,\dots,l_k}$. More importantly, changing $(a,q) \to (Ka,Kq)$ for any $K \in \mathbb{Z}_+$ does not affect the expression in (4.24).

Proposition 4.2. Let $\mathcal{P}: \mathbb{R} \to \mathbb{R}^k$ be a polynomial mapping whose components are polynomials with integer coefficients and different degrees. Let $N \in \mathbb{D}^B$, $M_1, \ldots, M_k \in \mathbb{R}_+$ and $l_1, \ldots, l_k \in \mathbb{N}$ with $l := \max\{l_1, \cdots, l_k\} \leq \tilde{c} \operatorname{Log}^{1/C_0} N$ with \tilde{c} given by (4.3). For each $\xi \in \mathbb{T}^k$, $\theta = (\theta_1, \ldots, \theta_k) \in \mathfrak{Q}_{l_1, \ldots, l_k}$ and $|\xi_i - \theta_i| \leq M_i^{-1}$ for all $i \in [k]$, we have

$$\left| m_{N,\Lambda_N}(\xi) - G^{\times}(\theta) \tilde{m}_{N,\mathbb{R}}(\xi - \theta) \right| \lesssim_{\mathcal{P}} 2^{kl} \exp(-c \log^{4/5} N) \left(1 + \max\{M_i^{-1} N^{d_i} : i \in [k]\} \right)$$
(4.26)

with c given as in (2.22).

Proof of Proposition 4.2. Denote $\theta = a/q$. Then q obeys (4.25) and $2^l \leq q \leq \exp(\operatorname{Log}^{1/C_0} N)$ (using the upper bound for l). By (2.22), for every $n \in J_N$ with $N \in \mathbb{D}^B$, we can deduce

$$\frac{S_n}{n} - \frac{\mathbb{1}_{(b,q)=1}}{\varphi(q)} =: a_n \quad \text{with} \quad |a_n| \lesssim \exp(-c \operatorname{Log}^{4/5} N), \tag{4.27}$$

where the sum S_n is given by

$$S_n := \sum_{j \in [n]} \Lambda_N(j) \mathbb{1}_{j \equiv b \pmod{q}}. \tag{4.28}$$

Using the congruence-based factorization, we have

$$m_{N,\Lambda_N}(\xi) = \sum_{b \in [q]} e(\theta \cdot \mathcal{P}(b)) E_{N,q,b}^{\mathcal{P}}(\xi - \theta),$$

where the average $E_{N,q,b}^{\mathcal{P}}$ is given by

$$E_{N,q,b}^{\mathcal{P}}(\zeta) := \frac{1}{\lfloor N \rfloor} \sum_{n \in J_N} e(\zeta \cdot \mathcal{P}(n)) \Lambda_N(n) \mathbb{1}_{n \equiv b \pmod{q}}, \qquad \zeta \in \mathbb{R}^k.$$
 (4.29)

Note $\Lambda_N(n) = 0$ unless (b,q) = 1. In fact, if $(b,q) \neq 1$, there exists an integer $r_0 \geq 2$ with $r_0|n$, which yields $\Lambda_N(n) = 0$. As a result, we have

$$m_{N,\Lambda_N}(\xi) = \sum_{b \in [q]^{\times}} e(\theta \cdot \mathcal{P}(b)) E_{N,q,b}^{\mathcal{P}}(\xi - \theta).$$

To prove (4.26), it suffices to show that for each $\zeta = (\zeta_1, \dots, \zeta_k) \in \mathbb{R}^k$ with $|\zeta_i| \leq M_i^{-1}$ $(i \in [k])$,

$$\left| E_{N,q,b}^{\mathcal{P}}(\zeta) - \frac{\mathbb{1}_{(b,q)=1}}{\varphi(q)} \tilde{m}_{N,\mathbb{R}}(\zeta) \right| \lesssim \exp(-c \operatorname{Log}^{4/5} N) \left(1 + \max\{M_i^{-1} N^{d_i} : i \in [k]\} \right)$$
(4.30)

with $E_{N,q,b}^{\mathcal{P}}$ given by (4.29). By combining (4.27)-(4.29), we rewrite the average $E_{N,q,b}^{\mathcal{P}}$ as

$$E_{N,q,b}^{\mathcal{P}}(\zeta) = \frac{1}{\lfloor N \rfloor} \sum_{n \in J_N} e(\zeta \cdot \mathcal{P}(n)) (S_n - S_{n-1})$$

$$= \frac{1}{\lfloor N \rfloor} \sum_{n \in J_N} e(\zeta \cdot \mathcal{P}(n)) \left(\frac{\mathbb{1}_{(b,q)=1}}{\varphi(q)} + b_n\right)$$
(4.31)

with $b_n = na_n - (n-1)a_{n-1}$. By the mean value theorem and a routine computation,

$$\left| \frac{1}{\lfloor N \rfloor} \sum_{n \in J_N} e(\zeta \cdot \mathcal{P}(n)) - \tilde{m}_{N,\mathbb{R}}(\zeta) \right| \leq \frac{1}{\lfloor N \rfloor} \sum_{n \in J_N} \left| e(\zeta \cdot \mathcal{P}(n)) - \int_{I_n} e(\zeta \cdot \mathcal{P}(t)) dt \right| + O(\frac{1}{N})$$

$$\lesssim N^{-1} \left(1 + \max\{M_i^{-1} N^{d_i} : i \in [k]\} \right),$$

where I_n denotes the interval [n, n+1]. This estimate immediately yields

$$\frac{\mathbb{1}_{(b,q)=1}}{\varphi(q)} \left| \frac{1}{\lfloor N \rfloor} \sum_{n \in I_N} e(\zeta \cdot \mathcal{P}(n)) - \tilde{m}_{N,\mathbb{R}}(\zeta) \right| \lesssim N^{-1} \left(1 + \max\{M_i^{-1} N^{d_i} : i \in [k]\} \right). \tag{4.32}$$

On the other hand, by using the Abel transform, the mean value theorem and the upper bound of a_n (as given in (4.27)), we have

$$\left| \frac{1}{\lfloor N \rfloor} \sum_{n \in J_N} e(\zeta \cdot \mathcal{P}(n)) b_n \right| \lesssim \frac{1}{\lfloor N \rfloor} \sum_{n \in J_N} n |a_n| \left| e(\zeta \cdot \mathcal{P}(n+1)) - e(\zeta \cdot \mathcal{P}(n)) \right| + \exp(-c \operatorname{Log}^{4/5} N)$$

$$\lesssim \exp(-c \operatorname{Log}^{4/5} N) \left(1 + \max\{M_i^{-1} N^{d_i} : i \in [k]\} \right),$$

which, along with (4.32) and (4.31), yields the desired (4.30).

In the following context, we use Proposition 4.2 to give a reduction of (4.20). Given functions $m: \mathbb{T}^k \to \mathbb{C}$ and $S: \mathbb{Q}^k \to \mathbb{C}$, we will be working with multilinear operators of the form

$$B^{l_1,\dots,l_k}[S;m](f_1,\dots,f_k)(x) := \sum_{\theta \in \mathfrak{Q}_{l_1,\dots,l_k}} S(\theta) \sum_{y \in \mathbb{Z}^k} K_{\tau_{\theta}m}(y) \prod_{i \in [k]} f_i(x-y_i)$$
(4.33)

for $x \in \mathbb{Z}$, where $\tau_{\theta} m(\xi) := m(\xi - \theta)$, and

$$K_m(y) := \int_{\mathbb{T}^k} m(\xi)e(-\xi \cdot y)d\xi. \tag{4.34}$$

Now we formulate our approximation result.

Proposition 4.3. Let $N \geq C_0$ with C_0 given by (4.5), and let $l_1, s_1, \ldots, l_k, s_k \in \mathbb{N}$ obeying $l_{(N)} \geq \max\{l, s\}$. Then, for all $f_1 \in \ell^{q_1}(\mathbb{Z}), \ldots, f_k \in \ell^{q_k}(\mathbb{Z})$ with $q_1, \ldots, q_k \in (1, \infty)$ such that $\frac{1}{q_1} + \cdots + \frac{1}{q_k} = \frac{1}{q} \leq 1$, we have

$$\|\tilde{A}_{N,\Lambda_{N}}(\tilde{\Pi}_{l_{1},s_{1}}^{N}f_{1},\ldots,\tilde{\Pi}_{l_{k},s_{k}}^{N}f_{k})-B^{l_{1},\ldots,l_{k}}[G^{\times};\tilde{m}_{N,\mathbb{R}}w_{N}^{s_{1},\ldots,s_{k}}](f_{1},\ldots,f_{k})\|_{\ell^{q}(\mathbb{Z})} \\ \lesssim \exp(-\frac{c}{2}\log^{4/5}N)\|f_{1}\|_{\ell^{q_{1}}(\mathbb{Z})}\cdots\|f_{k}\|_{\ell^{q_{k}}(\mathbb{Z})},$$

$$(4.35)$$

with c given as in (2.22), where $w_N^{s_1,...,s_k}(\xi) := \prod_{i \in [k]} \eta_N^{s_i}(\xi_i)$ with $\eta_N^{s_i}$ given by (4.17).

Proof of Proposition 4.3. Normalize $||f_i||_{\ell^{q_i}(\mathbb{Z})} = 1$ for all $i \in [k]$, and write

$$\tilde{A}_{N,\Lambda_N}(\tilde{\Pi}_{l_1,s_1}^N f_1,\ldots,\tilde{\Pi}_{l_k,s_k}^N f_k) = \sum_{\theta \in \mathfrak{Q}_{l_1,\ldots,l_k}} \sum_{y \in \mathbb{Z}^k} K_{m_{N,\Lambda_N} \times \tau_\theta w_N^{s_1,\ldots,s_k}}(y) \prod_{i \in [k]} f_i(x-y_i).$$

Applying Minkowski's inequality followed by Hölder's inequality, we bound the left-hand side of (4.35) by

$$\sum_{\theta \in \mathfrak{Q}_{l_1, \dots, l_k}} \| K_{\mathfrak{m}_{N, \theta, s}} \|_{\ell^1(\mathbb{Z}^k)} = \sum_{\theta \in \mathfrak{Q}_{l_1, \dots, l_k}} \| \int_{\mathbb{T}^k} \mathfrak{m}_{N, \theta, s}(\xi) \ e(-\xi \cdot y) d\xi \|_{\ell^1(y \in \mathbb{Z}^k)}, \tag{4.36}$$

where $K_{\mathfrak{m}_{N,\theta,s}}$ is defined by (4.34) with the function m replaced by

$$\mathfrak{m}_{N,\theta,s} := \left(m_{N,\Lambda_N} - G^{\times}(\theta) \tau_{\theta} \tilde{m}_{N,\mathbb{R}} \right) \tau_{\theta} w_N^{s_1,\dots,s_k}.$$

Invoking (2.18) and (3.3), we then deduce, through a routine computation, that for any tuples $(\beta_1, \ldots, \beta_k) \in \mathbb{N}^k$,

$$\|\partial_1^{\beta_1} \cdots \partial_k^{\beta_k} \mathfrak{m}_{N,\theta,s}\|_{L^{\infty}(\mathbb{T}^k)} \lesssim_{\beta_1,\dots,\beta_k} N^{\beta_1 d_1 + \dots + \beta_k d_k}. \tag{4.37}$$

Utilizing the support condition for $w_N^{s_1,\dots,s_k}$, integration by parts and (4.37), we can infer

$$|K_{\mathfrak{m}_{N,\theta,s}}(y)| \lesssim N^{-(d_1+\dots+d_k)} 2^{d_1s_1+\dots+d_ks_k} \prod_{i\in[k]} \langle y_i/N^{d_i} \rangle^{-4}.$$
 (4.38)

Furthermore, we have another bound for $K_{\mathfrak{m}_{N,\theta,s}}$ given by

$$|K_{\mathfrak{m}_{N,\theta,s}}(y)| \lesssim N^{-(d_1+\dots+d_k)} 2^{d_1s_1+\dots+d_ks_k} \|\mathfrak{m}_{N,\theta,s}\|_{L^{\infty}(\mathbb{T}^k)}.$$
 (4.39)

Taking the geometry average between (4.39) with (4.38), we obtain the bound

$$|K_{\mathfrak{m}_{N,\theta,s}}(y)| \lesssim 2^{kd_k s} \|\mathfrak{m}_{N,\theta,s}\|_{L^{\infty}(\mathbb{T}^k)}^{1/2} \prod_{i \in [k]} (N^{-d_i} \langle y_i / N^{d_i} \rangle^{-2}).$$
 (4.40)

On the other hand, it follows from Proposition 4.2 (with $M_i \sim N^{d_i} 2^{-d_i s_i}$) that

$$\|\mathfrak{m}_{N,\theta,s}\|_{L^{\infty}(\mathbb{T}^k)} \lesssim 2^{kl+d_k s} \exp(-c \operatorname{Log}^{4/5} N)$$
(4.41)

with c given as in (2.22). Inserting (4.41) into (4.40) gives a bound for $|K_{\mathfrak{m}_{N,\theta,s}}|$, which, combined with $l_{(N)} \geq \max\{l, s\}$ and (4.3), yields

$$(4.36) \lesssim \exp(-\frac{c}{2} \log^{4/5} N),$$
 as desired.

This ends the proof of Proposition 4.3.

By using Proposition 4.3, we now reduce the proof of (4.20) to showing

$$\| \left(B^{l_1, \dots, l_k} [G^{\times}; \tilde{m}_{N, \mathbb{R}} w_N^{s_1, \dots, s_k}] (f_1, \dots, f_k) \right)_{N \in \mathbb{D}_{l, s}} \|_{\ell^q(\mathbb{Z}; \mathbf{V}^r)}$$

$$\lesssim 2^{-c_0(l+s)} \sum_{j \in [k]} \left(\| f_j \|_{L^{q_{j_1}}(\mathbb{Z})} \prod_{i \in [k] \setminus \{j\}} \| f_i \|_{L^{q_{i_2}}(\mathbb{Z})} \right)$$
(4.42)

with $q, \{q_{jv}\}_{(j,v)\in[k]\times\{1,2\}}, r$ given as in Theorem 4.1. Once (4.42) is established, we can complete the proof of Theorem 1.3.

4.3.1. Model operators. For each $i \in [k]$, we denote

$$\eta_{Nt}^{s_i}(\zeta_i) := e(\zeta_i P_i(Nt)) \eta_N^{s_i}(\zeta_i), \quad \zeta = (\zeta_1, \dots, \zeta_k) \in \mathbb{R}^k, \tag{4.43}$$

where $\eta_N^{s_i}$ is defined by (4.17). Then we express $\tilde{m}_{N,\mathbb{R}} w_N^{s_1,\dots,s_k}$ as

$$\tilde{m}_{N,\mathbb{R}}(\zeta)w_N^{s_1,\dots,s_k}(\zeta) = \int_{1/2}^1 \left(\prod_{i\in[k]} \eta_{N,t}^{s_i}(\zeta_i)\right) dt.$$

Define a new quantity

$$u := \begin{cases} 100k(s^{5/4} + 1) & \text{if } s \ge C_* l, \\ 100k(l^{5/4} + 1) & \text{if } C_* l > s, \end{cases}$$
 (4.44)

where C_* fixed later is a sufficiently large constant.

Remark 5. The exponet 5/4, which can be substituted with any value in $(5/4, \infty)$, is determined by the error in (2.22). This is another reason why we cannot employ the metric entropy argument which plays a crucial role in [30].

We define the high-frequency case when $C_*l \leq s$, and the low-frequency case when $C_*l > s$. As we shall see later, the low-frequency case requires a more involved analysis. In fact, it is the low-frequency case that requires multi-frequency analysis techniques. For the high-frequency case, we will use the multilinear Weyl inequality in the continuous setting, as detailed in Appendix A. To obtain the desired estimate for the low-frequency case, we divide the set $\mathbb{D}_{l,s}$ into two subsets \mathbb{I}_{\leq} and $\mathbb{I}_{>}$ given by

$$\mathbb{I}_{\leq} := \mathbb{I}_{l,s} \leq \mathbb{D}_{l,s} \cap [1, 2^{10p_0 2^u}] \quad \text{and} \quad \mathbb{I}_{\geq} := \mathbb{I}_{l,s} \geq \mathbb{D}_{l,s} \cap (2^{10p_0 2^u}, \infty), \quad (4.45)$$

and reduce the matter to proving (4.42) with the set $\mathbb{D}_{l,s}$ replaced by \mathbb{I}_{\leq} and $\mathbb{I}_{>}$, which correspond to the low-frequency case at a small scale and a large scale, respectively. We will use the multilinear Rademacher-Menshov inequality and the harmonic analysis of the adelic integers $\mathbb{A}_{\mathbb{Z}}$ to address the small-scale case and the large-scale case, respectively.

Invoking the notation (2.8) and setting

$$F_{N,t}^{l_i,s_i} := F_{N,t}^{l_i,s_i}(f_i) := T_{\mathbb{Z}}^{\mathfrak{Q}_{l_i}}[\eta_{N,t}^{s_i}]f_i, \qquad i \in [k], \tag{4.46}$$

we may rewrite $B^{l_1,\dots,l_k}[G^{\times}(\theta);\tilde{m}_{N,\mathbb{R}}w_N^{s_1,\dots,s_k}](f_1,\dots,f_k)$ as

$$\int_{1/2}^{1} B^{l_1,\dots,l_k}[G^{\times};\eta_u^*](F_{N,t}^{l_1,s_1},\dots,F_{N,t}^{l_k,s_k})dt, \tag{4.47}$$

with the function η_u^* defined by

$$\eta_u^*(\zeta) = \eta_{\leq -10p_0u}(\zeta_1) \cdots \eta_{\leq -10p_0u}(\zeta_k), \quad \zeta = (\zeta_1, \dots, \zeta_k) \in \mathbb{R}^k,$$
 (4.48)

where p_0 is given as in (4.4), and u is given by (4.44). Note that $\eta_u^*(\zeta) = 1$ on the support $\prod_{i \in [k]} \eta_N^{s_i}(\zeta_i)$, since (4.3), (4.5) and (4.44). We provide a heuristic explanation for this process. In fact, this process is to move the parameter N from the multiplier of the multilinear operator $B^{l_1,\ldots,l_k}[G^{\times}(\theta); \tilde{m}_{N,\mathbb{R}}w_N^{s_1,\ldots,s_k}]$ to the functions f_1,\ldots,f_k , and the choice of η_u^* is utilized to establish (4.53) below. It now suffices to consider

$$B^{l_1,\dots,l_k}[G^{\times};\eta_u^*](F_{N,t}^{l_1,s_1},\dots,F_{N,t}^{l_k,s_k})$$
(4.49)

for any $t \in [1/2, 1]$. Repeating the arguments yielding (4.35) (or arguments in proving [31, Lemma 8.6]), we can deduce that the difference between (4.49) and the multilinear operator

$$\tilde{A}_{2^{u},\Lambda_{2^{u}}}\left(\Pi_{l_{1},\leq-10p_{0}u}(F_{N,t}^{l_{1},s_{1}}),\ldots,\Pi_{l_{k},\leq-10p_{0}u}(F_{N,t}^{l_{k},s_{k}})\right) \tag{4.50}$$

satisfies the error bound

$$\|\left((4.49) - (4.50)\right)_{N \in \mathbb{D}_{l,s}}\|_{\ell^{q}(\mathbb{Z}; \mathbf{V}^{r})}$$

$$\lesssim \exp\left(-\frac{c}{2}u^{4/5}\right) \prod_{i \in [k]} \|(F_{N,t}^{l_{i},s_{i}})_{N \in \mathbb{D}_{l,s}}\|_{\ell^{q_{i}}(\mathbf{V}^{r})} \lesssim 2^{-c(l+s)} \prod_{i \in [k]} \|f_{i}\|_{\ell^{q_{i}}(\mathbb{Z})}$$

$$(4.51)$$

for all $f_i \in \ell^{q_i}(\mathbb{Z})$ with $1 < q_1, \dots, q_k < \infty$ satisfying $1/q_1 + \dots + 1/q_k = 1/q \le 1$. The second inequality in (4.51) follows from

$$\| (F_{N,t}^{l_i,s_i})_{N \in \mathbb{D}_{l_s}} \|_{\ell^{q_i}(\mathbb{Z};\mathbf{V}^r)} \lesssim (s+1) (2^{\mathbf{C}_{q_i}(2^l)} \mathbb{1}_{l \ge 10} + 1) \|f_i\|_{\ell^{q_i}(\mathbb{Z})} \quad (i \in [k])$$

$$(4.52)$$

with⁸ $\mathbf{C}_{q_i}(\cdot)$ given as in (2.15). As a matter of fact, we can divide the proof of (4.52) into two cases: $s_i > 0$ and $s_i = 0$. For the case $s_i > 0$, we can use (2.14) and [31, Theorem B.1] to establish the stronger shifted square functions estimate resulting (4.52) via (2.6). As for the case $s_i = 0$, decompose $\eta_{N,t}^{s_i} = \eta_N^{s_i} + \mathcal{E}_{N,t}^i$. We establish the variational inequality (4.52) with $\eta_{N,t}^{s_i}$ replaced by $\eta_N^{s_i}$ by combining [30, Theorem 3.27] and the Lépingle's inequality [34, 39], while we can use the standard Littlewood-Paley arguments (see, for example, [14] or [42]) to control (4.52) with $\eta_{N,t}^{s_i}$ replaced by $\mathcal{E}_{N,t}^i$. Moreover, by Minkowski's inequality, (2.21), (4.46), (4.52) and Theorem 2.1, we can establish the following variational inequality:

$$\|((4.50))_{N\in\mathbb{D}_{l,s}}\|_{\ell^q(\mathbb{Z};\mathbf{V}^r)} \lesssim (s+1)^k (2^{\mathbf{C}_{q_1}(2^l)+\cdots+\mathbf{C}_{q_k}(2^l)} \mathbb{1}_{l\geq 10} + 1) \|f_1\|_{\ell^{q_1}(\mathbb{Z})} \cdots \|f_k\|_{\ell^{q_k}(\mathbb{Z})}$$
(4.53)

for all $f_i \in \ell^{q_i}(\mathbb{Z})$ with $1 < q_1, \dots, q_k < \infty$ satisfying $1/q_1 + \dots + 1/q_k = 1/q \le 1$.

4.3.2. Reduction of (4.42). Thanks to (4.53) and (4.51), we only need to consider

$$\max\{s, C_* l\} \ge 10. \tag{4.54}$$

We shall reduce the proof of (4.42) to showing the following propositions.

Proposition 4.4 (High-frequency case). Let $k \geq 2$, $s \geq C_* l$ with (4.54), and let $2 \leq q_1, \ldots, q_k < \infty$ such that $1/q_1 + \cdots + 1/q_k = 1$. Then there exists $c \in (0,1)$ independent of C_* such that, for all $f_1 \in \ell^{q_1}(\mathbb{Z}^k), \ldots, f_k \in \ell^{q_k}(\mathbb{Z}^k)$, we have

$$\|\sup_{N\in\mathbb{D}_{l,s}} |B^{l_1,\cdots,l_k}[G^{\times}; \tilde{m}_{N,\mathbb{R}} w_N^{s_1,\dots,s_k}](f_1,\cdots,f_k)|\|_{\ell^1(\mathbb{Z})} \lesssim 2^{-cs} \prod_{i\in[k]} \|f_i\|_{\ell^{q_i}(\mathbb{Z})}. \tag{4.55}$$

Proposition 4.5 (Low-frequency case at a small scale). Let $k \geq 2$, $C_*l > s$ with (4.54), and let $1 < q_1, \ldots, q_k < \infty$ such that $1/q_1 + \cdots + 1/q_k = 1/q \leq 1$. Then there exists $c_1 \in (0,1)$ such that for any $r \geq 2$ and for any $f_1 \in \ell^{q_1}(\mathbb{Z}^k), \ldots, f_k \in \ell^{q_k}(\mathbb{Z}^k)$, we have

$$\left\| \left(B^{l_1, \cdots, l_k} [G^{\times}; \eta_u^*] \left(F_{N,t}^{l_1, s_1}, \cdots, F_{N,t}^{l_k, s_k} \right) \right)_{N \in \mathbb{I}_{\leq}} \right\|_{\ell^q(\mathbb{Z}; \mathbf{V}^r)} \lesssim 2^{-c_1 l} \prod_{i \in [k]} \|f_i\|_{\ell^{q_i}(\mathbb{Z})}$$
(4.56)

for all $t \in [1/2, 1]$, where $\{F_{N,t}^{l_i, s_i}\}_{i \in [k]}$ and η_u^* are given by (4.46) and (4.48).

Proposition 4.6 (Low-frequency case at a large scale). Let $k \geq 2$ and $n \in \mathbb{N}$, let $I_n(d_k)$, q, q_* and $\{q_{jv}\}_{(j,v)\in[k]\times\{1,2\}}$ be given as in Theorem 4.1. Assume that $q_* < r < 2\rho_{d_k}^{n+1}$ and $C_*l > s$ with (4.54). Then there exists $c_2 \in (0,1)$ such that for any $f_j \in \cap_{v=1,2} \ell^{q_{jv}}(\mathbb{Z})$ with $j \in [k]$,

$$\left\| \left(B^{l_1, \dots, l_k} [G^{\times}, \eta_u^*] \left(F_{N, t}^{l_1, s_1}, \dots, F_{N, t}^{l_k, s_k} \right) \right)_{N \in \mathbb{I}_{>}} \right\|_{\ell^q(\mathbb{Z}; \mathbf{V}^r)} \\
\lesssim 2^{-c_2 l} \sum_{j \in [k]} \left(\| f_j \|_{L^{q_{j1}}(\mathbb{Z})} \prod_{i \in [k] \setminus \{j\}} \| f_i \|_{L^{q_{i2}}(\mathbb{Z})} \right), \tag{4.57}$$

⁸Evidently, we have $q_i \in [\tilde{p}'_0, \tilde{p}_0]$ for some sufficiently large $\tilde{p}_0 \in 2\mathbb{Z}_+$, where \tilde{p}_0 depends only on q_i . Thus, when applying Theorem 2.1, we do not need to distinguish between the constants \mathbf{C}_{q_i} and $\mathbf{C}_{\tilde{p}_0}$.

for all $t \in [1/2, 1]$, where $\{F_{N,t}^{l_i, s_i}\}_{i \in [k]}$ and η_u^* are given by (4.46) and (4.48).

Remark 6. As we shall see in the proof of (4.42) below, the restrictions on the functions f_1, \ldots, f_k and the parameters $q, \{q_{jv}\}_{(j,v)\in[k]\times\{1,2\}}$ in Theorem 4.1 are derived from the conditions in Proposition 4.6. Specifically, the restrictions arise from the applications of techniques leading to the arithmetic multilinear estimate (see Lemma 6.1). Moreover, if we consider the bilinear case (k=2) with special polynomials of the form (n,P(n)) obeying deg $P\geq 2$, we can expand the interval $I_n(d_k)$ ((1.10) plays a crucial role) and relax the lower bound of q, ultimately obtaining the same result as in [31] by setting $(q,q_{jv})=(1,2)$ for all $(j,v)\in\{1,2\}^2$ and using interpolation.

Proof of (4.42) (accepting Propositions 4.4-4.6). Using (4.51) and (4.53), we first deduce that for all $r \in (2, \infty)$,

$$\|((4.49))_{N\in\mathbb{D}_{l,s}}\|_{\ell^q(\mathbb{Z};\mathbf{V}^r)} \lesssim (s+1)^k (2^{\mathbf{C}_{p_0}(2^l)} \mathbb{1}_{l\geq 10} + 1) \prod_{i\in[k]} \|f_i\|_{\ell^{q_i}(\mathbb{Z})}$$
(4.58)

for all $q_1, \ldots, q_k \in [p'_0, p_0]$ with $i \in [k]$ such that $1/q_1 + \cdots + 1/q_k = 1/q \le 1$. By interpolating (4.58) with (4.55), and setting C_* sufficiently large, we can deduce (4.42) for the high-frequency case where $s \ge C_* l$. Moreover, inequality (4.56) in Proposition 4.5 implies (4.42) for the low-frequency case $(C_* l > s)$ at a small scale. Indeed, we have relaxed the constraints on f_1, \ldots, f_k and the parameters $q, \{q_{jv}\}_{(j,v)\in[k]\times\{1,2\}}$ in the above arguments. Finally, by invoking the definition of \mathbf{V}^r in Section 2 and applying Hölder's inequality, we can derive (4.42) for the low-frequency case $(C_* l > s)$ at a large scale from Proposition 4.6 and inequality (4.58).

We conclude this section by providing the proof of Proposition 4.4.

Proof of Proposition 4.4. Normalize $||f_i||_{\ell^{q_i}(\mathbb{Z})} = 1$ for all $i \in [k]$. We may reduce the matter to proving

$$\sup_{\theta \in \mathfrak{Q}_{l_1,\dots,l_k}} \|\sup_{N \in \mathbb{D}_{l,s}} |T_{\theta,N}(f_1,\dots,f_k)|\|_{\ell^1(\mathbb{Z})} \lesssim 2^{-c's}$$

$$(4.59)$$

for some c' > 0 independent of C_* , where the operator $T_{\theta,N}$ is given by

$$T_{\theta,N}(f_1,\ldots,f_k)(x) := \sum_{y \in \mathbb{Z}^k} K_{\tau_{\theta}(\tilde{m}_{N,\mathbb{R}}w_N^{s_1,\ldots,s_k})}(y) \prod_{i \in [k]} f_i(x-y_i), \quad x \in \mathbb{Z}.$$

Indeed, using Minkowski's inequality, the bound $\#\mathfrak{Q}_{l_1,\ldots,l_k} \leq 2^{2kl}$, and the notation (4.33), we can achieve (4.55) by setting C_* large enough such that $C_*c' > 100k$.

It remains to establish (4.59). For each $\theta \in \mathfrak{Q}_{l_1,\ldots,l_k}$, we observe the identity

$$T_{\theta,N}(f_1,\ldots,f_k)(x) = e((\theta_1+\cdots+\theta_k)x) \ T_{0,N}(M_{\theta_1}f_1,\ldots,M_{\theta_k}f_k)(x), \quad x \in \mathbb{Z},$$

where M_{θ_i} is the modulation operator defined by $M_{\theta_i}f(x) := e(-\theta_i x)f(x)$ for $i \in [k]$. In view of this identity, to achieve (4.59), it suffices to prove

$$\| \sup_{N \in \mathbb{D}_{l,s}} |T_{0,N}(g_1, \dots, g_k)| \|_{\ell^1(\mathbb{Z})} \lesssim 2^{-c's}$$
(4.60)

for any $g_i \in \ell^{q_i}(\mathbb{Z})$ such that $||g_i||_{\ell^{q_i}(\mathbb{Z})} = 1$ with $i \in [k]$. By employing the transference principle (see the arguments leading to [7, Lemma 4.4]), we can reduce the proof of (4.60) to demonstrating

$$\|\sup_{N\in\mathbb{D}_{l,s}} |\tilde{T}_{0,N}(g_1,\dots,g_k)|\|_{L^1(\mathbb{R})} \lesssim 2^{-c's}$$
(4.61)

for any $g_i \in L^{q_i}(\mathbb{R})$ such that $||g_i||_{L^{q_i}(\mathbb{R})} = 1$ for $i \in [k]$, where $\tilde{T}_{0,N}(g_1,\ldots,g_k)$ is defined by

$$\tilde{T}_{0,N}(g_1,\ldots,g_k)(x) := \int_{\mathbb{R}^k} K_{\tilde{m}_{N,\mathbb{R}} w_N^{s_1,\ldots,s_k}}(y) g_1(x-y_1) \cdots g_k(x-y_k) dy, \quad x \in \mathbb{R}.$$

We next prove (4.61). Write $\tilde{T}_{0,N}(g_1,\ldots,g_k)$ as

$$\int_{1/2}^{1} (\eta_N^{s_1}(D)g_1)(x - P_1(Nt)) \cdots (\eta_N^{s_k}(D)g_k)(x - P_k(Nt))dt, \tag{4.62}$$

where $\eta_N^{s_i}(D)g_i(x) := \int_{\mathbb{R}} \eta_N^{s_i}(\xi) \mathcal{F}_{\mathbb{R}}g_i(\xi)e(\xi x)d\xi$. Without loss of generality, we may assume that s is sufficiently large since (4.61) can be proved by Minkowski's inequality, Hölder's inequality and Young's inequality otherwise. In addition, we further assume

$$s_1 \le s_2 \le \ldots \le s_{k-1} \le s_k = s$$
 (4.63)

since other cases can be handled similarly. Set $s_0=0$ and let $n\in\mathbb{N}_{\leq k}$ be the biggest integer such that $s_n=0$. Obviously, $n\leq k-1$ (otherwise s=0). Without loss of generality, we only consider the case where $1\leq n\leq k$, as the case n=0 can be treated in the same way. For each $i\in[n]$ (so we have $s_i=0$), by Taylor's expansion $e(\xi P_i(t))=\sum_{m_i\in\mathbb{N}}(-2\pi\mathrm{i}\xi P_i(t))^{m_i}/(m_i!)$, we write $(\eta_N^{s_i}(D)g_i)(x-P_i(t))$ as

$$(\eta_N^{s_i}(D)g_i)(x - P_i(t)) = \int_{\mathbb{R}} \eta_N^0(\xi) \mathcal{F}_{\mathbb{R}} g_i(\xi) e(-\xi x) e(\xi P_i(t)) d\xi$$
$$= \sum_{m_i \in \mathbb{N}} \frac{(-2\pi i N^{-d_i} P_i(t))^{m_i}}{m_i!} \eta_N^{0, m_i}(D) g_i(x)$$

where $\eta_N^{0,m_i}(D)g_i(x) = \int_{\mathbb{R}} (\xi N^{d_i})^{m_i} \eta_N^0(\xi) \mathcal{F}_{\mathbb{R}} g_i(\xi) e(\xi x) d\xi$ can be seen as a variant of $\eta_N^0(D)g_i$. Rewrite (4.62) as

$$\sum_{m_1 \in \mathbb{N}} \cdots \sum_{m_n \in \mathbb{N}} \frac{(-2\pi i)^{m_1 + \dots + m_n}}{m_1! \cdots m_n!} \Big(\prod_{i \in [n]} \eta_N^{0, m_i}(D) g_i(x) \Big) \mathcal{A}_{N, \mathbb{R}}(g_{n+1}, \dots, g_k)(x), \tag{4.64}$$

where the operator $\mathcal{A}_{N,\mathbb{R}}$ is given by

$$\mathcal{A}_{N,\mathbb{R}}(g_{n+1},\ldots,g_k)(x) := \int_{1/2}^1 \left(\prod_{i=n+1}^k \left(\eta_N^{s_i}(D)g_i \right) (x - P_i(Nt)) \right) \mathcal{B}_N(t) dt$$

with $\mathcal{B}_N(t) := \prod_{i \in [n]} \left(\frac{P_i(Nt)}{N^{d_i}}\right)^{m_i}$ obeying $|\mathcal{B}_N| + |\mathcal{B}'_N| \lesssim \tilde{C}^{m_{[1,n]}}$ for some $\tilde{C} > 0$, where $m_{[1,n]} := m_1 + \cdots + m_n$. Define $\frac{1}{q_*} := \frac{1}{q_{n+1}} + \cdots + \frac{1}{q_k}$ and $\frac{1}{q_{**}} := \frac{1}{q_1} + \cdots + \frac{1}{q_n}$. This yields $\frac{1}{q_*} + \frac{1}{q_{**}} = 1$ and $q_*, q_{**} \in (1, \infty)$. By a routine computation, there exists C > 0 such that

$$\left\| \sup_{N \in \mathbb{D}_{l,s}} \left| \prod_{i \in [n]} \eta_N^{0,m_i}(D) g_i \right| \right\|_{L^{q_{**}}(\mathbb{R})} \lesssim C^{m_{[1,n]}} \prod_{i \in [n]} \|M_{HL} g_i\|_{L^{q_i}(\mathbb{R})} \lesssim C^{m_{[1,n]}} \prod_{i \in [n]} \|g_i\|_{L^{q_i}(\mathbb{R})}$$
(4.65)

with the implicit constant independent of m_1, \ldots, m_n , where M_{HL} represents the continuous Hardy-Littlewood maximal operator. By (4.65) and (4.64), to achieve (4.61), it suffices to show

$$\left\| \left(\sum_{N \in \mathbb{D}_{l,s}} \left| \mathcal{A}_{N,\mathbb{R}}(g_{n+1}, \dots, g_k) \right|^{q_*} \right)^{1/q_*} \right\|_{L^{q_*}(\mathbb{R})} \lesssim \tilde{C}^{m_{[1,n]}} 2^{-c's} \prod_{i=n+1}^k \|g_i\|_{L^{q_i}(\mathbb{R})}. \tag{4.66}$$

Invoking (4.63) (with $s = s_k > 0$) and applying Theorem A.1 for j = k (see Remark 8 below Theorem A.1), we bound the left-hand side of (4.66) by

$$\left(\sum_{N\in\mathbb{D}_{l,s}} \|\mathcal{A}_{N,\mathbb{R}}(g_{n+1},\ldots,g_{k})\|_{L^{q_{*}}(\mathbb{R})}^{q_{*}}\right)^{1/q_{*}} \lesssim \tilde{C}^{m_{[1,n]}} 2^{-c's} \prod_{i=n+1}^{k} \|\left(\eta_{N}^{s_{i}}(D)g_{i}\right)_{N\in\mathbb{D}_{l,s}}\|_{L^{q_{i}}(\mathbb{R};\ell^{q_{i}})} \\
\lesssim \tilde{C}^{m_{[1,n]}} 2^{-c's} \prod_{i=n+1}^{k} \|\left(\eta_{N}^{s_{i}}(D)g_{i}\right)_{N\in\mathbb{D}_{l,s}}\|_{L^{q_{i}}(\mathbb{R};\ell^{2})},$$

where the condition $q_{n+1}, \ldots, q_k \geq 2$ is used to the second inequality. This yields (4.66) by repeatedly applying standard Littlewood-Paley arguments (since $s_i > 0$ for all $i \in [k] \setminus [n]$). This concludes the proof of Proposition 4.4.

The proofs of Propositions 4.5 and 4.6 are postponed to the following two sections. Without loss of generality, we assume that l is large enough.

5. The low-frequency case at a small scale: Proof of Proposition 4.5

In this section, we prove Proposition 4.5 by establishing a multilinear Rademacher-Menshov inequality. A natural approach to derive this inequality would be to directly generalize the arguments used for the bilinear version in [31]. However, the techniques used in [31] grow substantially more intricate as the parameter k increases. Surprisingly, this obstacle can be circumvented. Specifically, by employing an inductive argument, the desired multilinear Rademacher-Menshov inequality can be deduced directly from the bilinear version established in [31].

Lemma 5.1 (Rademacher-Menshov for multilinear forms). Let $k \geq 2$, and $K \in \mathbb{Z}_+$. For each $i \in [k]$, let $\{f_{i,N}\}_{N \in [K]}$ be the elements of some vector space V_i . Let $0 < q < \infty$, and let $B: V_1 \times \cdots \times V_k \to L^q(X)$ be a multilinear map for some measure space X. Then

$$\| (B(f_{1,N}, \dots, f_{k,N}))_{N \in [K]} \|_{L^{q}(X; \mathbf{V}^{2})} \lesssim \langle \operatorname{Log} K \rangle^{k-2+k \max\{1, \frac{1}{q}\}} \times \sup \| B(\sum_{j \in [K]} \varepsilon_{j}^{(1)}(f_{1,j} - f_{1,j-1}), \dots, \sum_{j \in [K]} \varepsilon_{j}^{(k)}(f_{k,j} - f_{k,j-1})) \|_{L^{q}(X)}$$
(5.1)

with the conventions $f_{i,0} = 0$ for all $i \in [k]$, where the supremum is applied over

$$\varepsilon_i^{(i)} \in \{1, -1\} \qquad \qquad \text{for all} \quad (i, j) \in [k] \times [K].$$

Proof of Lemma 5.1. We achieve our objective using induction on the parameter k. It is important to highlight that the case k=2 was established in [31, Corollary 8.2]. For every $m \in \{3,4,5,\ldots\}$, we assume that (5.1) holds for k=m-1, and we will demonstrate that (5.1) also holds for k=m. For $(\tilde{n},n) \in [K]^2$, we denote

$$a_{\tilde{n},n} := B(f_{1,\tilde{n}}, \dots, f_{m-1,\tilde{n}}, f_{m,n}).$$
 (5.2)

By [31, Lemma 8.1] giving the two-dimensional Rademacher-Menshov inequality, we have

$$\|(a_{N,N})_{N\in[K]}\|_{\mathbf{V}^2} \lesssim \sum_{M_1,M_2\in 2^{\mathbb{N}}\cap[K]} \|(\Delta a_{M_1j_1,M_2j_2})_{(j_1,j_2)\in[K/M_1]\times[K/M_2]}\|_{\ell^2}, \tag{5.3}$$

where $\Delta a_{M_1j_1,M_2j_2} := a_{M_1j_1,M_2j_2} - a_{M_1j_1,M_2(j_2-1)} - a_{M_1(j_1-1),M_2j_2} + a_{M_1(j_1-1),M_2(j_2-1)}$ with $a_{k_1,k_2} = 0$ whenever $k_1k_2 = 0$. By (5.2), we expand the sum on the right-hand side of (5.3) as

$$\sum_{M_1, M_2 \in 2^{\mathbb{N}} \cap [K]} \left(\sum_{j_1 \in [K/M_1]} \sum_{j_2 \in [K/M_2]} \left| B(f_{1, M_1 j_1}, \dots, f_{m-1, M_1 j_1}, \tilde{f}_{m, M_2 j_2}) \right. \right)$$
(5.4)

$$-B(f_{1,M_1(j_1-1)},\ldots,f_{m-1,M_1(j_1-1)},\tilde{f}_{m,M_2j_2})|^2\Big)^{1/2}$$
 (5.5)

with $\tilde{f}_{m,M_2j_2} := f_{m,M_2j_2} - f_{m,M_2(j_2-1)}$. Denote

$$G_{m,K,M_2}(t) := \sum_{j_2 \in [K/M_2]} \varepsilon_{j_2}^{(m)}(t) \tilde{f}_{m,M_2 j_2}, \tag{5.6}$$

where $\{\varepsilon_i^{(m)}(t)\}_{i=0}^{\infty}$ is the sequence of Rademacher functions (see, e.g., [21, Appendix C]) on [0,1] with

$$\|\sum_{i=0}^{\infty} z_i \varepsilon_i^{(m)}(t)\|_{L_t^{\rho}([0,1])} \sim (\sum_{i=0}^{\infty} |z_i|^2)^{1/2} \quad \text{for any } \rho \in (0,\infty).$$
 (5.7)

Then, using (5.7) with $\rho = r \in (0, \min\{q, 2\})$, Minkowski's inequality and (5.6), we obtain from (5.5) that

$$\|(a_{N,N})_{N\in[K]}\|_{\mathbf{V}^{2}}$$

$$\lesssim \langle \operatorname{Log} K \rangle \sum_{M_{2}\in 2^{\mathbb{N}}\cap[K]} \| \left(B(f_{1,N},\ldots,f_{m-1,N},G_{m,K,M_{2}}(t)) \right)_{N\in[K]} \|_{L_{t}^{r}([0,1];\mathbf{V}^{2})}.$$

$$(5.8)$$

Taking the $L^q(X)$ norm on the both sides of (5.8), we deduce by Minkowski's inequality (for $q \ge 1$) or quasi-triangle inequality (for 0 < q < 1) that

$$\|(a_{N,N})_{N\in[K]}\|_{L^{q}(X;\mathbf{V}^{2})} \lesssim \langle \operatorname{Log} K \rangle^{1+\max\{1,\frac{1}{q}\}} \times \sup_{M_{2}\in2^{\mathbb{N}}\cap[K]} \sup_{t\in[0,1]} \|(B(f_{1,N},\cdots,f_{m-1,N},G_{m,K,M_{2}}(t)))_{N\in[K]}\|_{L^{q}(X;\mathbf{V}^{2})},$$
(5.9)

Now, by the induction assumption that (5.1) holds for k = m - 1, we conclude

$$\| \left(B(f_{1,N}, \cdots, f_{m-1,N}, G_{m,K,M_2}(t)) \right)_{N \in [K]} \|_{L^q(X; \mathbf{V}^2)} \lesssim \langle \operatorname{Log} K \rangle^{m-3+(m-1)\max\{1, \frac{1}{q}\}}$$

$$\times \sup \left\| B\left(\sum_{j \in [K]} \varepsilon_j^{(1)}(f_{1,j} - f_{1,j-1}), \cdots, \sum_{j \in [K]} \varepsilon_j^{(m-1)}(f_{m-1,j} - f_{m-1,j-1}), G_{m,K,M_2}(t) \right) \right\|_{L^q(X)}$$

for all $M_2 \in 2^{\mathbb{N}} \cap [K]$ and $t \in [0, 1]$, where the supremum is applied over $\varepsilon_j^{(i)} \in \{1, -1\}$ for all $(i, j) \in [m - 1] \times [K]$. This estimate, combined with (5.9) and (5.6), gives (5.1) for the case k = m. This completes the proof of Lemma 5.1.

Proof of Proposition 4.5. By (4.51), it suffices to show that for any $t \in [1/2, 1]$,

$$\|((4.50))_{N \in \mathbb{I}_{<}}\|_{\ell^{q}(\mathbb{Z}; \mathbf{V}^{r})} \lesssim 2^{-cl} \|f_{1}\|_{\ell^{q_{1}}(\mathbb{Z})} \cdots \|f_{k}\|_{\ell^{q_{k}}(\mathbb{Z})}$$
(5.10)

with r, q, q_1, \ldots, q_k given as in Proposition 4.5.

Normalize $||f_i||_{\ell^{q_i}(\mathbb{Z})} = 1$ for all $i \in [k]$. We can enumerate the elements of \mathbb{I}_{\leq} in order as $N_1 < \cdots < N_K$, which, with (4.45), gives $K = O(2^u)$. By Lemma 5.1, we can reduce the matter to showing the inequality

$$u^{O(1)} \| \tilde{A}_{2^{u}, \Lambda_{2^{u}}} (\Pi_{l_{1}, \leq -10p_{0}u}(\tilde{F}_{N,t}^{l_{1}, s_{1}}), \cdots, \Pi_{l_{k}, \leq -10p_{0}u}(\tilde{F}_{N,t}^{l_{k}, s_{k}})) \|_{\ell^{q}(\mathbb{Z})} \lesssim 2^{-cl}, \tag{5.11}$$

with the functions $\tilde{F}_{N,t}^{l_i,s_i} \ (i \in [k])$ given by

$$\tilde{F}_{N,t}^{l_i,s_i} := \tilde{F}_{N,t}^{l_i,s_i}(f_i) := T_{\mathbb{Z}}^{\mathfrak{Q}_{l_i}}[\eta_{*,N,t}^{s_i}]f_i, \quad \text{where} \quad \eta_{*,N,t}^{s_i} = \sum_{j \in [k]} \varepsilon_j^{(i)}(\eta_{N_j,t}^{s_i} - \eta_{N_{j-1},t}^{s_i})$$

with $\{\varepsilon_j^{(i)}\}_{(i,j)\in[k]\times[K]}\subset\{1,-1\}$. Since $\mathcal{F}_{\mathbb{Z}}\tilde{F}_{N,t}^{l_i,s_i}$ vanishes on $\mathfrak{M}_{\leq l_i-1,\leq -d_i\operatorname{Log}N+d_il_i-d_i}$ for each $i\in[k]$, we deduce by Theorem 3.1 and Theorem 2.1 that

$$\|\tilde{A}_{2^{u},\Lambda_{2^{u}}}(\Pi_{l_{1},\leq-10p_{0}u}(\tilde{F}_{N,t}^{l_{1},s_{1}}),\cdots,\Pi_{l_{k},\leq-10p_{0}u}(\tilde{F}_{N,t}^{l_{k},s_{k}}))\|_{\ell^{q}(\mathbb{Z})} \lesssim 2^{-cl} \prod_{i\in[k]} \|\tilde{F}_{N,t}^{l_{i},s_{i}}\|_{\ell^{q_{i}}(\mathbb{Z})}.$$
 (5.12)

By the shifted Calderón-Zygmund theory (see [31, Theorem B.1]) and (2.14), we have

$$\|\tilde{F}_{N,t}^{l_i,s_i}\|_{\ell^{q_i}(\mathbb{Z})} \lesssim \langle s \rangle^{O(1)} 2^{\mathbf{C}_{q_i}(2^l)}, \qquad i \in [k]. \tag{5.13}$$

Since $u \sim l^{5/4}$, $C_* l > s$ and l is sufficiently large, we can obtain (5.11) by combining (5.13) and (5.12).

6. The low-frequency case at a large scale: Proof of Proposition 4.6

In this section, we shall prove Proposition 4.6 by employing harmonic analysis of the adelic integers $\mathbb{A}_{\mathbb{Z}}$. More precisely, we will establish a general arithmetic multilinear estimate along with some crucial p-adic estimates. Interestingly, these p-adic estimates apply only to certain linear averages, yet they are sufficient for the proof.

6.1. Reduction of Proposition 4.6. We first write (4.47) in another useful way. Since $N \in \mathbb{I}_{>}$, we replace the function η_n^* in (4.57) by

$$\eta_u^{**}(\zeta) = \eta_{\leq -p_0 2^u}(\zeta_1) \cdots \eta_{\leq -p_0 2^u}(\zeta_k), \quad \zeta = (\zeta_1, \dots, \zeta_k) \in \mathbb{R}^k.$$

This process is necessary and crucial for applying the quantitative Shannon sampling theorem (see Theorem B.1 below). Introduce the adelic model functions $f_{l_1,\mathbb{A}},\ldots,f_{l_k,\mathbb{A}}$ by the formula

$$f_{l_i,\mathbb{A}}(x,y) = \sum_{\alpha_i \in (\mathbb{Q}/\mathbb{Z})_{l_i}} \int_{\mathbb{R}} \eta_{\leq -p_0 2^u}(\xi_i) \mathcal{F}_{\mathbb{Z}} f_i(\alpha_i + \xi_i) e(-(\xi_i, \alpha_i) \cdot (x, y)) d\xi_i, \qquad i \in [k]$$

where $(x,y) \in \mathbb{A}_{\mathbb{Z}} = \mathbb{R} \times \hat{\mathbb{Z}}$, or equivalently on the Fourier side

$$\mathcal{F}_{\mathbb{A}_{\mathbb{Z}}} f_{l_i,\mathbb{A}}(\xi_i,\alpha_i) = \mathbb{1}_{h(\alpha_i)=2^{l_i}} \eta_{\leq -p_0 2^u}(\xi_i) \mathcal{F}_{\mathbb{Z}} f_i(\alpha_i + \xi_i), \qquad i \in [k],$$

where $\xi_i \in \mathbb{R}$ and $\alpha_i \in \mathbb{Q}/\mathbb{Z}$. For each $i \in [k]$, by (4.44), one can also interpret $f_{l_i,\mathbb{A}}$ as the interpolated functions

$$f_{l_i,\mathbb{A}} = \mathcal{S}_{\mathbb{R}_{\leq -p_0 2^u \times (\mathbb{Q}/\mathbb{Z})_{l_i}}^{-1}} \Pi_{l_i, \leq -p_0 2^u} f_i$$

and $\Pi_{l_i, \leq -p_0 2^u} f_i = \mathcal{S} f_{l_i, \mathbb{A}}$, where the operators \mathcal{S} and $\mathcal{S}_{\Omega}^{-1}$ are given by (B.3) and (B.4) respectively. This, with Theorem B.1 and (2.14), gives that for each $q \in (1, \infty)$,

$$||f_{l_i,\mathbb{A}}||_{L^q(\mathbb{A}_{\mathbb{Z}})} \lesssim 2^{\mathbf{C}_q(2^l)} ||f_i||_{\ell^q(\mathbb{Z})} \tag{6.1}$$

with $C_q(\cdot)$ given by (2.15). Similarly, invoking (2.7), (2.9) and the notation in Subsection 2.4, we can write (4.47) (with η_u^{**} in place of η_u^*) as

$$\mathcal{S}\left(\int_{1/2}^{1} B_{\mathbb{A}_{\mathbb{Z}}}[1 \otimes m_{l,\hat{\mathbb{Z}}^{\times}}](\mathfrak{F}_{N,t,\mathbb{A}}^{l_{1},s_{1}}, \cdots, \mathfrak{F}_{N,t,\mathbb{A}}^{l_{k},s_{k}})dt\right),\tag{6.2}$$

with $\mathfrak{F}_{N,t,\mathbb{A}}^{l_i,s_i}$ and $m_{l,\hat{\mathbb{Z}}^{\times}}$ given by

$$\mathfrak{F}_{N,t,\mathbb{A}}^{l_{i},s_{i}}(x,y) := \mathfrak{F}_{N,t}^{s_{i}}(f_{l_{i},\mathbb{A}})(x,y) := T_{\mathbb{A}_{\mathbb{Z}}}[\eta_{N,t}^{s_{i}} \otimes 1]f_{l_{i},\mathbb{A}}(x,y) \text{ and }$$
 (6.3)

$$m_{l,\hat{\mathbb{Z}}^{\times}}(\theta) := m_{l,\hat{\mathbb{Z}}^{\times}}(\theta_1, \dots, \theta_k) := G^{\times}(\theta) \prod_{i \in [k]} \mathbb{1}_{(\mathbb{Q}/\mathbb{Z})_{l_i}}(\theta_i).$$

$$(6.4)$$

Because N is only related to the continuous part \mathbb{R} in the adelic group, for any r > 2 and $q \in (1, \infty)$, we have

$$\|\left(\mathfrak{F}_{N,t,\mathbb{A}}^{l_i,s_i}\right)_{N\in\mathbb{I}_{\sim}}\|_{L^q(\mathbb{A}_{\mathbb{Z}};\mathbf{V}^r)} \lesssim \langle s \rangle^{O(1)} \|f_{l_i,\mathbb{A}}\|_{L^q(\mathbb{A}_{\mathbb{Z}})}. \tag{6.5}$$

In fact, similar to the proof of (4.52), we will divide the proof of (6.5) into two cases: $s_i > 0$ and $s_i = 0$. For the case $s_i > 0$, we can use [31, Theorem B.1] to establish the stronger shifted square functions estimate yielding (6.5) since (2.5). For the case $s_i = 0$, we can achieve (6.5) using Lépingle's inequality and standard Littlewood-Paley arguments.

Since $N > 2^{10p_0 2^u}$, the above functions $(f_{l_1,\mathbb{A}}, \dots, f_{l_k,\mathbb{A}})$ and $(\mathfrak{F}_{N,t,\mathbb{A}}^{l_1,s_1}, \dots, \mathfrak{F}_{N,t,\mathbb{A}}^{l_k,s_k})$ defined on $\mathbb{A}_{\mathbb{Z}}$ have Fourier support in the region

$$\left(\mathbb{R}_{\leq -2^u} \times (\mathbb{Q}/\mathbb{Z})_{l_1}\right) \times \cdots \times \left(\mathbb{R}_{\leq -2^u} \times (\mathbb{Q}/\mathbb{Z})_{l_k}\right)$$

which is contained in

$$\left(\mathbb{R}_{\leq -2^{u}} \times \left(\frac{1}{Q_{< l}} \mathbb{Z}/\mathbb{Z}\right)\right) \times \dots \times \left(\mathbb{R}_{\leq -2^{u}} \times \left(\frac{1}{Q_{< l}} \mathbb{Z}/\mathbb{Z}\right)\right) \tag{6.6}$$

with $Q_{\leq l} := \text{lcm}(q \in \mathbb{Z}_+ : q \in [2^l]) \leq 2^{O(2^l)}$ $(l = \max\{l_1, \dots, l_k\})$, which is obviously smaller than $2^{2^{u/4}}$ since (4.44). Thus, in this large-scale situation, we deduce by applying Theorem B.1 (with the regime (6.6) and the normed vector space $B = \mathbf{V}^r$) that

$$\|\left((6.2)\right)_{N\in\mathbb{I}_{>}}\|_{\ell^{q}(\mathbb{Z};\mathbf{V}^{r})}\sim\|\left(\int_{1/2}^{1}B_{\mathbb{A}_{\mathbb{Z}}}[1\otimes m_{l,\hat{\mathbb{Z}}^{\times}}](\mathfrak{F}_{N,t,\mathbb{A}}^{l_{1},s_{1}},\cdots,\mathfrak{F}_{N,t,\mathbb{A}}^{l_{k},s_{k}})dt\right)_{N\in\mathbb{I}_{>}}\|_{\ell^{q}(\mathbb{A}_{\mathbb{Z}};\mathbf{V}^{r})}.$$

This, combined with inequality (6.1), yields that it is sufficient for Proposition 4.6 to demonstrate that there is c > 0 such that for any $t \in [1/2, 1]$,

$$\|\left(B_{\mathbb{A}_{\mathbb{Z}}}\left[1\otimes m_{l,\hat{\mathbb{Z}}^{\times}}\right]\left(\mathfrak{F}_{N,t,\mathbb{A}}^{l_{1},s_{1}},\cdots,\mathfrak{F}_{N,t,\mathbb{A}}^{l_{k},s_{k}}\right)\right)_{N\in\mathbb{I}_{>}}\|_{\ell^{q}(\mathbb{A}_{\mathbb{Z}};\mathbf{V}^{r})}$$

$$\lesssim 2^{-cl}\sum_{j\in[k]}\left(\|f_{l_{j},\mathbb{A}}\|_{L^{q_{j_{1}}}(\mathbb{A}_{\mathbb{Z}})}\prod_{i\in[k]\setminus\{j\}}\|f_{l_{i},\mathbb{A}}\|_{L^{q_{i_{2}}}(\mathbb{A}_{\mathbb{Z}})}\right),\tag{6.7}$$

where the parameters $r, q, \{q_{jv}\}_{(j,v)\in[k]\times\{1,2\}}$ are given as in Proposition 4.6.

6.2. **Proof of (6.7).** Invoke the notation (2.9) and (6.4), and then rewrite the multilinear operator $B_{\mathbb{A}_{\mathbb{Z}}}[1 \otimes m_{L\hat{\mathbb{Z}}^{\times}}]$ in (6.2) as

$$\sum_{\theta \in (\mathbb{Q}/\mathbb{Z})^k} \int_{\xi \in \mathbb{R}^k} G^{\times}(\theta) \Big(\prod_{i \in [k]} \mathcal{F}_{\mathbb{A}_{\mathbb{Z}}} \mathfrak{F}_{N,t,\mathbb{A}}^{l_i,s_i}(\xi_i,\theta_i) \Big) e\Big(- (x,y) \cdot (\xi_1 + \dots + \xi_k, \theta_1 + \dots + \theta_k) \Big) d\xi \quad (6.8)$$

with $\theta = (\theta_1, \dots, \theta_k)$ and $\xi = (\xi_1, \dots, \xi_k)$. By the Fourier inverse transform in the continuous setting, (6.8) equals

$$\sum_{\theta_1 \in (\mathbb{Q}/\mathbb{Z})_{l_1}} \cdots \sum_{\theta_k \in (\mathbb{Q}/\mathbb{Z})_{l_k}} G^{\times}(\theta) \Big(\prod_{i \in [k]} \mathcal{F}_{\hat{\mathbb{Z}}} \mathfrak{F}_{N,t,\mathbb{A}}^{l_i,s_i}(x,\theta_i) \Big) e\Big(-y(\theta_1 + \cdots + \theta_k) \Big). \tag{6.9}$$

Let $A_{\hat{\mathbb{Z}}^{\times}}$ be the multilinear operator acting on functions defined over the profinite integers $\hat{\mathbb{Z}}$. Precisely, for any positive integer Q, and any functions $g_1, \ldots, g_k : \mathbb{Z}/Q\mathbb{Z} \to \mathbb{C}$ (which one can also view as functions on $\hat{\mathbb{Z}}$ in the obvious fashion), the operator $A_{\hat{\mathbb{Z}}^{\times}}$ is defined by the formula

$$A_{\hat{\mathbb{Z}}^{\times}}(g_1,\ldots,g_k)(y) := \mathbb{E}_{n \in (\mathbb{Z}/Q\mathbb{Z})^{\times}} g_1(y - P_1(n)) \cdots g_k(y - P_k(n)).$$

As stated previously, changing $(a,q) \to (Ka, Kq)$ for any $K \in \mathbb{Z}_+$ does not affect the expression (4.24) of $G^{\times}(\frac{a}{q})$. Thus, by expanding the notation of G^{\times} and using the Fourier inverse transform in the arithmetic setting, to bound (6.9), it suffices to estimate the function $A_{\hat{\mathbb{Z}}^{\times}}(\mathfrak{F}_{N,t,\mathbb{A}}^{l_1,s_1}(x,\cdot),\ldots,\mathfrak{F}_{N,t,\mathbb{A}}^{l_k,s_k}(x,\cdot))(y)$. This allows us to reduce the proof of (6.7) to proving that there exists $R_0 \in (2,r)$ such that for any $t \in [1/2,1]$ and $x \in \mathbb{R}$,

$$\|\left(A_{\widehat{\mathbb{Z}}^{\times}}(\mathfrak{F}_{N,t,\mathbb{A}}^{l_{1},s_{1},x},\ldots,\mathfrak{F}_{N,t,\mathbb{A}}^{l_{k},s_{k},x})\right)_{N\in\mathbb{I}_{>}}\|_{L^{q}(\widehat{\mathbb{Z}};\mathbf{V}^{r})}$$

$$\lesssim 2^{-cl}\sum_{j\in[k]}\left(\|\left(\mathfrak{F}_{N,t,\mathbb{A}}^{l_{j},s_{j},x}\right)_{N\in\mathbb{I}_{>}}\|_{L^{q_{j_{1}}}(\widehat{\mathbb{Z}};\mathbf{V}^{R_{0}})}\prod_{i\in[k]\setminus\{j\}}\|\left(\mathfrak{F}_{N,t,\mathbb{A}}^{l_{i},s_{i},x}\right)_{N\in\mathbb{I}_{>}}\|_{L^{q_{i_{2}}}(\widehat{\mathbb{Z}};\mathbf{V}^{R_{0}})}\right)$$

$$(6.10)$$

with $r, q, \{q_{jv}\}_{(j,v)\in[k]\times\{1,2\}}$ given as in Proposition 4.6, where $\mathfrak{F}_{N,t,\mathbb{A}}^{l_k,s_k,x}(y) := \mathfrak{F}_{N,t,\mathbb{A}}^{l_k,s_k}(x,y)$ with $y\in\hat{\mathbb{Z}}$, and the implicit constant is independent of x and t. Indeed, once (6.10) holds, we can

achieve (6.7) by taking $L^q(x \in \mathbb{R})$ norm on both sides of (6.10), using Hölder's inequality and (6.5) (with $r = \mathbb{R}_{\circ}$).

We introduce the following lemma. Invoking the notation $\rho_{\rm d}$ in Subsection 4.1, we denote

$$S_n := \{ (q, r) \in \mathbb{R}^2 : 2\rho_{d_k}^n \le q < r < 2\rho_{d_k}^{n+1} \}, \quad n \in \mathbb{N}.$$
 (6.11)

Lemma 6.1 (Arithmetic multilinear estimate). Let $k \geq 2$, $n \in \mathbb{N}$, and $1 \in \mathbb{N}$. Assume that $(q,r) \in S_n$, and $1 < q_1, \dots, q_k < \infty$ such that $\frac{1}{q_1} + \dots + \frac{1}{q_k} = \frac{1}{q}$. Fix $j \in [k]$. Then we have

$$||A_{\hat{\mathbb{Z}}^{\times}}(g_1, \cdots, g_k)||_{L^r(\hat{\mathbb{Z}})} \lesssim 2^{-cl} ||g_1||_{L^{q_1}(\hat{\mathbb{Z}})} \cdots ||g_k||_{L^{q_k}(\hat{\mathbb{Z}})}$$
(6.12)

for any $g_1 \in L^{q_1}(\hat{\mathbb{Z}}), \ldots, g_k \in L^{q_k}(\hat{\mathbb{Z}})$ with $\mathcal{F}_{\hat{\mathbb{Z}}}g_i$ vanishing on $(\mathbb{Q}/\mathbb{Z})_{\leq 1}$.

Remark 7. We achieve (6.12) for all $k \geq 2$ without imposing any restrictions on the polynomial map \mathcal{P} beyond those specified in (3.1) and (3.2). However, we need the condition (6.11) to achieve the norm interchanging (see (6.13) below) and the crucial p-adic estimates (see (6.26) and (6.27) below).

We now prove (6.10) by accepting Lemma 6.1. Invoking the norm interchanging trick from [31, Lemma 9.5], for any $1 \le R_1 < r_1 \le \infty$, we have the inequality

$$||(f_k)_{k\in[K]}||_{L^{r_1}(X;\mathbf{V}^{r_1})} \lesssim_{r_1,R_1} ||(f_k)_{k\in[K]}||_{\mathbf{V}^{R_1}([K];L^{r_1}(X))}, \tag{6.13}$$

where X is a measure space. Invoke that $q, q_*, \{q_{jv}\}_{(j,v) \in [k] \times \{1,2\}}$ and r are given as in Proposition 4.6. So we have $(q,r) \in S_n$ (given by (6.11)). Let $R \in (2,r)$ fixed later be an absolute constant. Using Hölder's inequality (since $q \leq q_* < r$) and (6.13) with $(R_1, r_1) = (R, r)$, we bound the left-hand side of (6.10) by

$$\|\left(A_{\hat{\mathbb{Z}}^{\times}}(\mathfrak{F}_{N,t,\mathbb{A}}^{l_{1},s_{1},x},\ldots,\mathfrak{F}_{N,t,\mathbb{A}}^{l_{k},s_{k},x})\right)_{N\in\mathbb{I}_{>}}\|_{L^{r}(\hat{\mathbb{Z}};\mathbf{V}^{r})} \lesssim \|\left(A_{\hat{\mathbb{Z}}^{\times}}(\mathfrak{F}_{N,t,\mathbb{A}}^{l_{1},s_{1},x},\ldots,\mathfrak{F}_{N,t,\mathbb{A}}^{l_{k},s_{k},x})\right)_{N\in\mathbb{I}_{>}}\|_{\mathbf{V}^{R}(\mathbb{I}_{>};L^{r}(\hat{\mathbb{Z}}))}.$$

$$(6.14)$$

Invoke the definitions (2.3) and (2.4). To bound the right-hand side of (6.14), it suffices to bound the following two expressions: first,

$$\sup \left(\sum_{n \in [J-1]} \| A_{\hat{\mathbb{Z}}^{\times}}(\mathfrak{F}_{N_{n+1},t,\mathbb{A}}^{l_1,s_1,x}, \dots, \mathfrak{F}_{N_{n+1},t,\mathbb{A}}^{l_k,s_k,x}) - A_{\hat{\mathbb{Z}}^{\times}}(\mathfrak{F}_{N_n,t,\mathbb{A}}^{l_1,s_1,x}, \dots, \mathfrak{F}_{N_n,t,\mathbb{A}}^{l_k,s_k,x}) \|_{L^r(\hat{\mathbb{Z}})}^R \right)^{1/R}, \quad (6.15)$$

where the supremum is applied over $J \in \mathbb{Z}_+$ and the sequence $\{N_n\}_{n \in \mathbb{N}_{\leq J}} \subset \mathbb{I}_{\geq}$; and second,

$$\|A_{\hat{\mathbb{Z}}^{\times}}(\mathfrak{F}_{N_{0},t,\mathbb{A}}^{l_{1},s_{1},x},\ldots,\mathfrak{F}_{N_{0},t,\mathbb{A}}^{l_{k},s_{k},x})\|_{L^{r}(\hat{\mathbb{Z}})},\tag{6.16}$$

with $N_0 := 2^{10p_0 2^{2u}} \in \mathbb{I}_{>}$. Because the desired estimate for (6.16) is a direct result of Lemma 6.1, it remains to estimate (6.15).

6.1, it remains to estimate (6.15). Since $\mathcal{F}_{\hat{\mathbb{Z}}} \mathfrak{F}_{N_n,t,\mathbb{A}}^{l_i,s_i,x}$ and $\mathcal{F}_{\hat{\mathbb{Z}}} \mathfrak{F}_{N_{n+1},t,\mathbb{A}}^{l_i,s_i,x}$ are supported on $(\mathbb{Q}/\mathbb{Z})_{l_i}$ for each $i \in [k]$, by Lemma 6.1 and a routine computation, (6.15) is

$$\lesssim 2^{-cl} \sum_{j \in [k]} \left\{ \sup \left(\sum_{n \in [J-1]} \|\bar{\mathfrak{F}}_{N_n,t,\mathbb{A}}^{l_j,s_j,x}\|_{L^{q_{j1}}(\hat{\mathbb{Z}})}^R \right)^{1/R} \left(\prod_{i \in [k] \setminus \{j\}} \|\sup_{N \in \mathbb{I}_{>}} |\mathfrak{F}_{N,t,\mathbb{A}}^{l_i,s_i,x}|\|_{L^{q_{i2}}(\hat{\mathbb{Z}})} \right) \right\}$$
(6.17)

with $\bar{\mathfrak{F}}_{N_n,t,\mathbb{A}}^{l_j,s_j,x}:=\mathfrak{F}_{N_{n+1},t,\mathbb{A}}^{l_j,s_j,x}-\mathfrak{F}_{N_n,t,\mathbb{A}}^{l_j,s_j,x}$. Set R such that $R\in(q_*,r)$. Applying Minkowski's inequality along with (6.17), we obtain

$$(6.15) \lesssim 2^{-cl} \sum_{j \in [k]} \Big\{ \| \big(\mathfrak{F}_{N,t,\mathbb{A}}^{l_j,s_j,x} \big)_{N \in \mathbb{I}_>} \|_{L^{q_{j1}}(\hat{\mathbb{Z}};\mathbf{V}^R)} \big(\prod_{i \in [k] \backslash \{j\}} \| \sup_{N \in \mathbb{I}_>} \| \mathfrak{F}_{N,t,\mathbb{A}}^{l_i,s_i,x} | \|_{L^{q_{i2}}(\hat{\mathbb{Z}})} \big) \Big\},$$

which can be bounded by the right side of (6.10) with $R_{\circ} = R$ (since (2.4)). This ends the proof of (6.10) with $R_{\circ} = R$, and we complete the proof of Proposition 4.6.

6.3. **Proof of Lemma 6.1.** By interpolation, we reduce the proof of Lemma 6.1 to establishing two key estimates. The first estimate states that for each $1/q_1 + \cdots + 1/q_k = 1/q \le 1$ with $1 < q_1, \ldots, q_k < \infty$,

$$||A_{\hat{\mathbb{Z}}^{\times}}(g_1, \cdots, g_k)||_{L^q(\hat{\mathbb{Z}})} \lesssim 2^{-cl} ||g_1||_{L^{q_1}(\hat{\mathbb{Z}})} \cdots ||g_k||_{L^{q_k}(\hat{\mathbb{Z}})}$$
(6.18)

for any $g_1 \in L^{q_1}(\hat{\mathbb{Z}}), \ldots, g_k \in L^{q_k}(\hat{\mathbb{Z}})$ such that $\mathcal{F}_{\hat{\mathbb{Z}}}g_j$ vanishing on $(\mathbb{Q}/\mathbb{Z})_{\leq 1}$, while the second estimate states that

$$||A_{\hat{\mathbb{Z}}^{\times}}(g_1, \cdots, g_k)||_{L^r(\hat{\mathbb{Z}})} \lesssim ||g_1||_{L^{q_1}(\hat{\mathbb{Z}})} \cdots ||g_k||_{L^{q_k}(\hat{\mathbb{Z}})}$$
(6.19)

with the parameters r, q_1, \ldots, q_k given as in Lemma 6.1. Here, the frequency support condition $\mathcal{F}_{\hat{\mathbb{Z}}}g_j = 0$ on $(\mathbb{Q}/\mathbb{Z})_{\leq 1}$ applies only to the first estimate.

Proof of (6.18). The proof adapts the approach from [31] to fit our weighted operator setting. By Minkowski's inequality and Hölder's inequality, we can establish (6.18) without the factor 2^{-cl} . Consequently, through interpolation, it suffices to prove (6.18) for the special case where $(q, q_1, \ldots, q_k) = (1, k, \ldots, k)$. By a standard limiting argument, we may assume that the functions $\{g_i\}_{i \in [k]}$ on $\hat{\mathbb{Z}}$ are periodic modulo Q, that is, they factor through $\mathbb{Z}/Q\mathbb{Z}$. The matter therefore reduces to establishing the bound

$$||A_{(\mathbb{Z}/Q\mathbb{Z})^{\times}}(g_1, \cdots, g_k)||_{L^1(\mathbb{Z}/Q\mathbb{Z})} \lesssim 2^{-cl} ||g_1||_{L^k(\mathbb{Z}/Q\mathbb{Z})} \cdots ||g_k||_{L^k(\mathbb{Z}/Q\mathbb{Z})},$$
 (6.20)

under the assumption that $\mathcal{F}_{\mathbb{Z}/Q\mathbb{Z}}g_j$ vanishes on $(\mathbb{Q}/\mathbb{Z})_{\leq 1} \cap (\frac{1}{Q}\mathbb{Z}/\mathbb{Z})$.

Let N and R be two large natural numbers, which should be considered as being much larger than l and Q. Define $g_{1,R}, \ldots, g_{k,R} \in \mathbf{S}(\mathbb{Z})$ by

$$g_{i,R}(n) := R^{-1/k} \phi(n/R) g_i(n \mod Q), \qquad i \in [k],$$

where $\phi \in S(\mathbb{R})$ is a non-nagetive and even function with $\int_{\mathbb{R}} \phi^k(x) dx = 1$ and the Fourier support is on [-1,1]. One has $g_{i,R} \in L^2(\mathbb{Z}) \cap L^k(\mathbb{Z})$ and $\mathcal{F}_{\mathbb{Z}}g_{j,R}$ is supported on $\pi([-1/R,1/R] \times \{\alpha \in \frac{1}{Q}\mathbb{Z}/\mathbb{Z} : \mathcal{F}_{\mathbb{Z}/Q\mathbb{Z}}g_j(\alpha) \neq 0\})$ (see (B.1) in the Appendix B for the definition of the map π). So $\mathcal{F}_{\mathbb{Z}}g_{j,R}$ vanishes on $\mathfrak{M}_{\leq l,\leq -d_i \log N+d_i l}$, which, with Theorem 3.1, gives

$$||A_{N,\Lambda_N}(g_{1,R},\cdots,g_{k,R})||_{\ell^1(\mathbb{Z})} \lesssim 2^{-cl}||g_{1,R}||_{\ell^k(\mathbb{Z})}\cdots||g_{k,R}||_{\ell^k(\mathbb{Z})}.$$
(6.21)

Normalize $||g_i||_{L^k(\mathbb{Z}/Q\mathbb{Z})} = 1$ for all $i \in [k]$. By a routine computation and the Riemann integrability of ϕ^k , we infer

$$\lim_{R \to \infty} \|g_{i,R}\|_{\ell^k(\mathbb{Z})} = \|g_i\|_{L^k(\mathbb{Z}/Q\mathbb{Z})} = 1, \qquad i \in [k].$$

This, combined with (6.21), leads to

$$\limsup_{N \to \infty} \limsup_{R \to \infty} ||A_{N,\Lambda_N}(g_{1,R}, \cdots, g_{k,R})||_{\ell^1(\mathbb{Z})} \lesssim 2^{-c!}.$$
(6.22)

By the congruence-based factorization, we rewrite $A_{N,\Lambda_N}(g_{1,R},\cdots,g_{k,R})(x)$ as

$$\sum_{a \in [Q]} \left(\prod_{i \in [k]} g_i(x \bmod Q - P_i(a)) \right) \mathbb{E}_{n \in [N]} \left(\Lambda_N(n) \mathbb{1}_{n \equiv a \pmod Q} \left\{ R^{-1} \prod_{i \in [k]} \phi(\frac{x - P_i(n)}{R}) \right\} \right). \tag{6.23}$$

By the mean value theorem (for each function $\phi(\cdot/R)$) and (2.22), we rewrite (6.23) as

$$R^{-1}\phi^{k}(x/R)\left\{A_{(\mathbb{Z}/Q\mathbb{Z})^{\times}}(g_{1},\ldots,g_{k})(x \bmod Q) + O(Q^{O(1)}\exp(-c\log^{4/5}N)\right\} + O(N^{O(1)}R^{-2}\langle x/R\rangle^{-2}).$$
(6.24)

(We do not aim to refine the error terms in (6.24), as the current estimates are sufficient.) Taking the $\ell^1(x \in \mathbb{Z})$ norm on both sides of (6.24), we deduce by using the congruence-based factorization, the Riemann integrability of ϕ^k and a routine computation that

$$\limsup_{N\to\infty} \limsup_{R\to\infty} \|A_{N,\Lambda_N}(g_{1,R},\ldots,g_{k,R})\|_{\ell^1(\mathbb{Z})} = \|A_{(\mathbb{Z}/Q\mathbb{Z})^{\times}}(g_1,\ldots,g_k)\|_{L^1(\mathbb{Z}/Q\mathbb{Z})},$$

which, together with (6.22), leads to the desired (6.20). This finishes the proof of (6.18).

Proof of (6.19). Denote

$$I_m := (1, \rho_{d_m}) \qquad (m \in [k])$$

with the notation ρ_{d_m} defined as $\rho_{\mathrm{d}}|_{\mathrm{d}=d_m}$ (see Subsection 4.1). We **claim** that for any $s \in I_k$ and for any $\tau \geq 1$,

$$\begin{cases}
\|A_{\hat{\mathbb{Z}}^{\times}}\|_{L^{2\tau}(\hat{\mathbb{Z}})\times L^{\infty}(\hat{\mathbb{Z}})\times \cdots \times L^{\infty}(\hat{\mathbb{Z}}) \to L^{2\tau s}(\hat{\mathbb{Z}})} \lesssim 1, \\
\|A_{\hat{\mathbb{Z}}^{\times}}\|_{L^{\infty}(\hat{\mathbb{Z}})\times L^{2\tau}(\hat{\mathbb{Z}})\times \cdots \times L^{\infty}(\hat{\mathbb{Z}}) \to L^{2\tau s}(\hat{\mathbb{Z}})} \lesssim 1, \\
\vdots \\
\|A_{\hat{\mathbb{Z}}^{\times}}\|_{L^{\infty}(\hat{\mathbb{Z}})\times L^{\infty}(\hat{\mathbb{Z}})\times \cdots \times L^{2\tau}(\hat{\mathbb{Z}}) \to L^{2\tau s}(\hat{\mathbb{Z}})} \lesssim 1.
\end{cases}$$
(6.25)

We first go ahead by accepting this claim. For every $s \in I_k$, by interpolation and $L^{r_1}(\hat{\mathbb{Z}}) \subset L^{r_2}(\hat{\mathbb{Z}})$ whenever $r_1 \geq r_2$, we can obtain from (6.25) (for $\tau = \rho_{d_k}^n$) that (6.19) holds for $r = 2\rho_{d_k}^n s$ and for any $\frac{1}{q_1} + \cdots + \frac{1}{q_k} = \frac{1}{q} \in [0, \frac{1}{2\rho_{d_k}^n}]$. In particular, we can achieve (6.19).

We next prove the above-mentioned claim. By Hölder's inequality, we may reduce the goal to showing the estimate for the linear average $A_{\hat{\mathbb{Z}}^{\times}}^{P_j}$, that is, $\|A_{\hat{\mathbb{Z}}^{\times}}^{P_j}\|_{L^2(\hat{\mathbb{Z}}) \to L^{2s}(\hat{\mathbb{Z}})} \lesssim 1$ for every $s \in I_j = (1, \rho_{d_j})$. By approximating $\hat{\mathbb{Z}}$ (and their unit group $\hat{\mathbb{Z}}^{\times}$) by the product of finitely many of the p-adic groups \mathbb{Z}_p (and \mathbb{Z}_p^{\times} respectively), and following the procedure in [31, p.1084], the problem reduces to establishing two key p-adic estimates for the averaging operator $A_{\mathbb{Z}_p^{\times}}^{P_j}$ defined by $A_{\mathbb{Z}_p^{\times}}^{P_j}(g) := \mathbb{E}_{n \in \mathbb{Z}_p^{\times}} g(x - P_j(n))$: first,

$$||A_{\mathbb{Z}_p^{\times}}^{P_j}||_{L^2(\mathbb{Z}_p)\to L^{2s}(\mathbb{Z}_p)} \lesssim_s \sqrt{\frac{p}{p-1}}.$$

$$(6.26)$$

for all primes p, and second, the improved uniform bound

$$\|A_{\mathbb{Z}_{p}^{s}}^{P_{j}}\|_{L^{2}(\mathbb{Z}_{p})\to L^{2s}(\mathbb{Z}_{p})} \le 1$$
 (6.27)

for all sufficiently large primes p. We will use the basic inequality

$$|A_{\mathbb{Z}_p^{\times}}^{P_j}(g)| \le \frac{p}{p-1} A_{\mathbb{Z}_p}^{P_j}(|g|),$$
 (6.28)

with the operator $A_{\mathbb{Z}_p}^{P_j}$ defined by $A_{\mathbb{Z}_p}^{P_j}(g) := \mathbb{E}_{n \in \mathbb{Z}_p} g(x - P_j(n))$. The estimate (6.27) is crucial, since the product $\prod_{p \in \mathbb{P}} \frac{p}{p-1}$ diverges.

We begin by proving (6.26). Our proof builds upon the counting bound from [31, Corollary C.2], which gives that for any $j \in \mathbb{N}$ and any prime p, the $L^{s_d}(\mathbb{Z}/p^j\mathbb{Z})$ norm of

$$h(m) = \# \left\{ n \in \mathbb{Z}/p^j\mathbb{Z}: \ P(n) = m \right\}$$

with the polynomial P of degree d is $O_{s_d}(1)$ (independent of p, j) for all $s_d \in (1, \frac{d}{d-1})$ if $d \geq 2$, and $s_d \in (1, \infty]$ if d = 1 (this case is easier since P is a linear polynomial). Thus, by Minkowski's inequality and a limit procedure, we have

$$||A_{\mathbb{Z}_p}^{P_j}(|g|)||_{L^s(\mathbb{Z}_p)} \lesssim ||g||_{L^1(\mathbb{Z}_p)}$$

for all $s \in I_i = (1, \rho_{d_i})$. This inequality, with (6.28), yields

$$\|A_{\mathbb{Z}_p^{\times}}^{P_j}(g)\|_{L^{2s}(\mathbb{Z}_p)}^2 \leq \|A_{\mathbb{Z}_p^{\times}}^{P_j}(|g|^2)\|_{L^{s}(\mathbb{Z}_p)} \leq \frac{p}{p-1}\|A_{\mathbb{Z}_p}^{P_j}(|g|^2)\|_{L^{s}(\mathbb{Z}_p)} \lesssim \frac{p}{p-1}\|g\|_{L^2(\mathbb{Z}_p)}^2.$$

Taking the square root on both sides of the above inequality yields the desired (6.26). We now prove (6.27). It suffices to show that for any sufficiently large prime p,

$$||A_{\mathbb{Z}_p^{\kappa}}^{P_j}(g)||_{L^{2s}(\mathbb{Z}_p)} \le ||g||_{L^2(\mathbb{Z}_p)}.$$
 (6.29)

We normalize $||g||_{L^2(\mathbb{Z}_p)} = 1$. Set $a := \mathbb{E}_{n \in \mathbb{Z}_p} g$, and $g_0 := g - a$. A simple computation gives that $\mathbb{E}_{n \in \mathbb{Z}_p} g_0(n) = 0$ and $|a| \le ||g||_{L^2(\mathbb{Z}_p)} = 1$. Performing a similar procedure yielding (6.18), we can infer that there exists $\gamma_0 > 0$ such that

$$||A_{\mathbb{Z}_{p}^{n}}^{P_{j}}(g_{0})||_{L^{2}(\mathbb{Z}_{p})} \lesssim p^{-\gamma_{0}}||g_{0}||_{L^{2}(\mathbb{Z}_{p})}.$$
 (6.30)

Indeed, the proof of (6.30) (dealing with the linear operator $A_{\mathbb{Z}_p^{\times}}^{P_j}$) is easier than that of (6.18). In addition, by interpolating (6.30) with (6.26), we can deduce that for each $s \in I_j$, there exists $\gamma_s > 0$ such that

$$||A_{\mathbb{Z}_{p}^{n}}^{P_{j}}(g_{0})||_{L^{2s}(\mathbb{Z}_{p})} \lesssim_{s} p^{-\gamma_{s}} ||g_{0}||_{L^{2}(\mathbb{Z}_{p})}.$$

$$(6.31)$$

Denote $B := \|g_0\|_{L^2(\mathbb{Z}_p)}^2$. Then we have $B \leq 1$ and $|a| = (1-B)^{1/2}$. Note that $A_{\mathbb{Z}_p^{\times}}^{P_j}(g) = a + A_{\mathbb{Z}_p^{\times}}^{P_j}(g_0)$, and the function $x \mapsto |x|^{2s}$ (with s > 1) is continuously twice differentiable. By Taylor's expansion, we then have

$$|A_{\mathbb{Z}_p^{p_i}}^{P_j}(g)|^{2s} = |a|^{2s} + 2s|a|^{2s-1}A_{\mathbb{Z}_p^{p_i}}^{P_j}(g_0) + O_s(|A_{\mathbb{Z}_p^{p_i}}^{P_j}(g_0)|^2 + |A_{\mathbb{Z}_p^{p_i}}^{P_j}(g_0)|^{2s}),$$

which, with $\mathbb{E}_{n\in\mathbb{Z}_n}g_0=0$ and (6.28), yields

$$||A_{\mathbb{Z}_{p}^{s}}^{P_{j}}(g)||_{L^{2s}(\mathbb{Z}_{p})}^{2s} \leq 1 - B + O_{s}(||A_{\mathbb{Z}_{p}}^{P_{j}}(g_{0})||_{L^{2}(\mathbb{Z}_{p})}^{2} + ||A_{\mathbb{Z}_{p}}^{P_{j}}(g_{0})||_{L^{2s}(\mathbb{Z}_{p})}^{2s}).$$
(6.32)

Substituting (6.31) into (6.32) gives that there exist two constants $C_s > 0$ and $\beta_s > 0$ such that

$$||A_{\mathbb{Z}_{n}^{N}}^{P_{j}}(g)||_{L^{2s}(\mathbb{Z}_{p})}^{2s} \le 1 - B + C_{s}p^{-\beta_{s}}B.$$

$$(6.33)$$

Set the prime p large enough such that $C_s p^{-\beta_s} \leq 1/2$. Then (6.27) follows from (6.33).

APPENDIX A. MULTILINEAR WEYL INEQUALITY IN THE CONTINUOUS SETTING

In this section, we will show the multilinear Weyl inequality in the continuous setting. For $k \in \mathbb{Z}_+$, we define a multilinear operator in the continuous setting

$$\tilde{A}_{N;\mathbb{R}}^{\mathcal{P}}(f_1,\dots,f_k)(x) := \frac{1}{N} \int_{N/2}^{N} f_1(x - P_1(t)) \cdots f_k(x - P_k(t)) dt, \qquad x \in \mathbb{R},$$
(A.1)

where $\mathcal{P} := (P_1, \dots, P_k)$ be a polynomial mapping with real coefficients and (3.2). Below we will use $\langle F, G \rangle_{\mathbb{R}^d}$ or $\langle F(y), G(y) \rangle_{y \in \mathbb{R}^d}$ to denote the inner product of the functions F and G in the continuous setting \mathbb{R}^d for $d \in \mathbb{Z}_+$. Let $\{e_i\}_{i \in [k]}$ be the standard unit vectors in \mathbb{R}^k whenever k > 2.

Theorem A.1. Let $N \geq 1$, $l \in \mathbb{N}$, $k \in \mathbb{Z}_+$, and let \mathcal{P} be given as in (A.1). Let $1 < q_1, \ldots, q_k < \infty$ be exponents such that $\frac{1}{q_1} + \cdots + \frac{1}{q_k} = \frac{1}{q} \leq 1$. There exists a small $c \in (0,1)$, possibly depending on $k, \mathcal{P}, q, q_1, \ldots, q_k$, such that the following holds. Let $f_i \in L^{q_i}(\mathbb{R})$ for

 $i \in [k]$, and fix $j \in [k]$. If $f_j \in L^{q_j}(\mathbb{R}) \cap L^2(\mathbb{R})$ and $\mathcal{F}_{\mathbb{R}} f_j$ vanishes on $[-N^{-d_j} 2^{d_j l}, N^{-d_j} 2^{d_j l}]$, then

$$\|\tilde{A}_{N;\mathbb{R}}^{\mathcal{P}}(f_1,\ldots,f_k)\|_{L^q(\mathbb{R})} \le c^{-1}(2^{-cl}+N^{-c})\|f_1\|_{L^{q_1}(\mathbb{R})}\cdots\|f_k\|_{L^{q_k}(\mathbb{R})}.$$
(A.2)

Remark 8. The interval [N/2, N] in (A.1) can be replaced by $[c_1N, c_2N]$ for any $c_2 > c_1 \ge 0$. Moreover, similar conclusion also holds for the general operator $T_{N:\mathbb{R}}^{\mathcal{P}}$ given by

$$T_{N;\mathbb{R}}^{\mathcal{P}}(f_1,\ldots,f_k)(x) := \int_{1/2}^1 f_1(x-P_1(Nt))\cdots f_k(x-P_k(Nt))\rho_N(t)dt,$$

where the function ρ_N satisfies the bound $|\rho'_N(t)| + |\rho_N(t)| \lesssim 1$ for all $1/2 \leq t \leq 1$.

Proof. By following the arguments in $Remark\ 2$ (see Section 3), we can reduce the proof of (A.2) to demonstrating

$$\|\tilde{A}_{N:\mathbb{R}}^{\mathcal{P}}(f_1,\dots,f_k)\|_{L^q(\mathbb{R})} \lesssim 2^{-cl} \|f_1\|_{L^{q_1}(\mathbb{R})} \cdots \|f_k\|_{L^{q_k}(\mathbb{R})}$$
(A.3)

for sufficiently large l satisfying $2^l \leq N$, where the implicit constant depends on $k, \mathcal{P}, q, q_1, \ldots, q_k$. Note that (A.3) without the decay 2^{-cl} follows from Minkowski's inequality and Hölder's inequality. The linear case (that is, k=1) can easily be proved by establishing the desired L^2 bound (which can be obtained by Plancherel's identity and elementary bounds for some basic oscillatory integrals) and using interpolation. We therefore restrict our attention to the case $k \geq 2$. By interpolation, it suffices to prove (A.3) for the case q=1. We next combine a transference trick with [30, Theorem 6.1] to achieve this goal.

Normalize $||f_i||_{L^{q_i}(\mathbb{R})} = 1$ for each $i \in [k]$. Let φ denote a smooth and nonnegative function on \mathbb{R}^{k-1} , which satisfies $\mathbb{1}_{[-1/2,1/2]^{k-1}} \leq \varphi \leq \mathbb{1}_{[-1,1]^{k-1}}$. Let h_{\circ} be any given function in $L^{\infty}(\mathbb{R})$, let $\epsilon > 0$ be a sufficiently small constant (which we will send to 0). Define

$$h(y) := h_{\circ}(y_1 + \dots + y_k), \qquad y = (y_1, \dots, y_k) \in \mathbb{R}^k,$$

$$g_{i,\epsilon}(y) := f_i(y_1 + \dots + y_k) \epsilon^{\frac{k-1}{q_i}} \varphi(\epsilon \ y_{\{j\}^c}), \quad i \in [k].$$
(A.4)

with the notation $y_{\{j\}^c} := (y)_{\{j\}^c} \in \mathbb{R}^{k-1}$ given by the vector $y \in \mathbb{R}^k$ without the j-th component. Obviously, by the change of variables and $\int \varphi^{q_i} \sim_{q_i,k} 1$, we have $g_{j,\epsilon} \in L^2(\mathbb{R}^k)$ and

$$||h||_{L^{\infty}(\mathbb{R}^k)} = ||h_{\circ}||_{L^{\infty}(\mathbb{R})}, \qquad ||g_{i,\epsilon}||_{L^{q_i}(\mathbb{R}^k)} \sim ||f_i||_{L^{q_i}(\mathbb{R})} = 1, \quad i \in [k].$$
 (A.5)

From these results with the support condition for $\mathcal{F}_{\mathbb{R}}f_j$ we obtain that $g_{1,\epsilon}, \ldots, g_{k,\epsilon}$ obey all conditions in [30, Theorem 6.1] (for $\mathbb{K} = \mathbb{R}$). Thus, by [30, Theorem 6.1] and the L^{q_i} bound for $g_{i,\epsilon}$ in (A.5), there exists $0 < c_1 < 1$, depending on $k, \mathcal{P}, q, q_1, \ldots, q_k$, such that

$$\left| \left\langle \frac{1}{N} \int_{N/2}^{N} \prod_{i \in [k]} g_{i,\epsilon}(y - P_i(t)e_i) dt, h(y) \right\rangle_{y \in \mathbb{R}^k} \right| \le c_1^{-1} 2^{-c_1 l} ||h||_{L^{\infty}(\mathbb{R}^k)}. \tag{A.6}$$

By (A.4) and the change of variables $y_j \to x - \sum_{i \in [k] \setminus \{j\}} y_i$, we can rewrite the inner product on the left-hand side of (A.6) as

$$\left\langle \frac{1}{N} \int_{N/2}^{N} \bar{\Phi}_{\epsilon}(t) f_{1}(x - P_{1}(t)) \cdots f_{k}(x - P_{k}(t)) dt, h_{\circ}(x) \right\rangle_{x \in \mathbb{R}}$$
(A.7)

where the function Φ_{ϵ} is given by

$$\bar{\Phi}_{\epsilon}(t) := \int_{y_{\{j\}^c} \in \mathbb{R}^{k-1}} \Phi_{\epsilon}(y_{\{j\}^c}, t) dy_{\{j\}^c} \quad \text{with} \quad \Phi_{\epsilon}(y_{\{j\}^c}, t) := \epsilon^{k-1} \prod_{i \in [k]} \varphi \left(\epsilon \ (y - P_i(t)e_i)_{\{j\}^c}\right).$$

Combining (A.5), (A.7) and (A.6), we can get

$$|(A.7)| \lesssim 2^{-cl} ||h_{\circ}||_{L^{\infty}(\mathbb{R})}. \tag{A.8}$$

Taking the limit as $\epsilon \to 0$ on both sides of (A.8), and then applying the Lebesgue dominant convergence theorem, we obtain

$$|\langle \tilde{A}_{N:\mathbb{R}}^{\mathcal{P}}(f_1,\ldots,f_k),h_{\circ}\rangle_{\mathbb{R}}|\lesssim 2^{-cl}||h_{\circ}||_{L^{\infty}(\mathbb{R})},$$

since $\lim_{\epsilon \to 0} \bar{\Phi}_{\epsilon}(t) = \int \varphi^k \sim_k 1$ for any $t \in [N/2, N]$. This completes the proof of (A.3) by using the dual arguments.

APPENDIX B. SAMPLING MAP AND QUANTITATIVE SHANNON SAMPLING THEOREM

In this section, we introduce several important maps utilized in the proof of the major arcs estimates. For motivations and detailed relationships, we refer to [31, Section 4]. The inclusion homomorphism $\iota: \mathbb{Z} \to \mathbb{A}_{\mathbb{Z}}$ is defined by $\iota(x) := (x, ((x \mod p^j)_{j \in \mathbb{N}})_{p \in \mathbb{P}})$ and the addition homomorphism $\pi: \mathbb{R} \times \mathbb{Q}/\mathbb{Z} \to \mathbb{T}$ is given by

$$\pi(\theta, \alpha) := \alpha + \theta; \tag{B.1}$$

these two maps are Fourier adjoint to each other in the sense that

$$\iota(x) \cdot \xi = x \cdot \pi(\xi) \tag{B.2}$$

for all $x \in \mathbb{Z}$ and $\xi \in \mathbb{R} \times \mathbb{Q}/\mathbb{Z}$. In the major arcs regimes, which is non-aliasing (the map π is injective on this type of regimes), we use these homomorphisms to "approximate" \mathbb{Z} by $\mathbb{A}_{\mathbb{Z}}$, which in principle decouples the discrete harmonic analysis of \mathbb{Z} from the continuous harmonic analysis of \mathbb{R} and the arithmetic harmonic analysis of \mathbb{Z} .

The inclusion homomorphism ι mentioned above leads to a sampling map $\mathcal{S}: \mathbf{S}(\mathbb{A}_{\mathbb{Z}}) \to \mathbf{S}(\mathbb{Z})$, which is defined as follows:

$$\mathcal{S}f(x) := f(\iota(x)) \tag{B.3}$$

for $x \in \mathbb{Z}$ and $f \in \mathbf{S}(\mathbb{A}_{\mathbb{Z}})$. Dually, the addition homomorphism $\pi : \mathbb{R} \times \mathbb{Q}/\mathbb{Z} \to \mathbb{T}$ leads to a projection map $\mathcal{T} : \mathbf{S}(\mathbb{R} \times \mathbb{Q}/\mathbb{Z}) \to \mathbf{S}(\mathbb{T})$ given by

$$\mathcal{T}F(\xi) := \sum_{(\theta,\alpha)\in\pi^{-1}(\xi)} F(\theta,\alpha)$$

for $\theta \in \mathbb{R}$, $\alpha \in \mathbb{Q}/\mathbb{Z}$, and $F \in S(\mathbb{R} \times \mathbb{Q}/\mathbb{Z})$. It is important to note that the definition of $S(\mathbb{R} \times \mathbb{Q}/\mathbb{Z})$ ensures that this sum contains at most countably many non-zero terms. We obtain from (B.2) the identity $\mathcal{F}_{\mathbb{Z}}^{-1} \circ \mathcal{T} = \mathcal{S} \circ \mathcal{F}_{\mathbb{A}_{\mathbb{Z}}}^{-1}$ or equivalently $\mathcal{F}_{\mathbb{Z}} \circ \mathcal{S} = \mathcal{T} \circ \mathcal{F}_{\mathbb{A}_{\mathbb{Z}}}$. For a non-aliasing compact set of adelic frequencies Ω , we define an interpolation operator

$$\mathcal{S}_{\Omega}^{-1} \colon \boldsymbol{S}(\mathbb{Z})^{\pi(\Omega)} \to \boldsymbol{S}(\mathbb{A}_{\mathbb{Z}})^{\Omega},$$
 (B.4)

where $S(\mathcal{G})^{\Omega}$ denotes the subspace of $S(\mathcal{G})$ consisting of functions that are Fourier supported on Ω for any LCA group \mathcal{G} . This operator, together with the sampling operator \mathcal{S} , forms a unitary correspondence between $\ell^2(\mathbb{Z})^{\pi(\Omega)}$ and $L^2(\mathbb{A}_{\mathbb{Z}})^{\Omega}$, as shown by Plancherel's theorem and the commutative diagram (given by [31, (4.6)] with \mathcal{P} replaced by \mathcal{T}).

Finally, we invoke the quantitative Shannon sampling theorem (see [31, Theorem 4.18]).

Theorem B.1. Let $0 < q \le \infty$, and B be a finite-dimensional normed vector space. If $F \in S(\mathbb{A}_{\mathbb{Z}}; B)$ has Fourier support in $[-\frac{c_0}{Q}, \frac{c_0}{Q}] \times \frac{1}{Q}\mathbb{Z}/\mathbb{Z}$ for some $Q \in \mathbb{Z}_+$ and some $0 < c_0 < 1/2$, then we have

$$\|\mathcal{S}F\|_{\ell^q(\mathbb{Z};B)} \sim_{c_0,q} \|F\|_{L^q(\mathbb{A}_{\mathbb{Z}};B)} \tag{B.5}$$

where we extend the sampling operator S to vector-valued functions in the obvious fashion.

References

- [1] T. Austin, On the norm convergence of non-conventional ergodic averages, Ergodic Theory Dynam. Systems **30** (2010), pp. 321–338.
- [2] V. Bergelson, Ergodic Ramsey Theory an update, Ergodic Theory of \mathbb{Z}^d -actions (edited by M. Pollicott and K. Schmidt), London Math. Soc. Lecture Note Series **228** (1996), pp. 1–61.
- [3] V. Bergelson, A. Leibman, A nilpotent Roth theorem, Invent. Math. 147 (2002), pp. 429-470.
- [4] V. Bergelson, Combinatorial and Diophantine Applications of Ergodic Theory (with appendices by A. Leibman and by A. Quas and M. Wierdl), Handbook of Dynamical Systems, Vol. 1B, B. Hasselblatt and A. Katok, eds., Elsevier (2006), pp. 745–841.
- [5] G. Birkhoff, Proof of the ergodic theorem, Proc. Natl. Acad. Sci. USA 17 (1931), no. 12, pp. 656-660.
- [6] J. Bourgain, Double recurrence and almost sure convergence, J. Reine Angew. Math. 404 (1990), pp. 140–161.
- [7] J. Bourgain, Pointwise ergodic theorems for arithmetic sets, Inst. Hautes Études Sci. Publ. Math. 69 (1989), 5-45
- [8] J. Bourgain, An approach to pointwise ergodic theorems, Geometric aspects of functional analysis (1986/87),
 Lecture Notes in Math., 317, 204–223. Springer, Berlin, 1988.
- [9] J. Bourgain, M. Mirek, E.M. Stein, B. Wróbel, Dimension-free estimates for discrete Hardy-Littlewood averaging operators over the cubes in \mathbb{Z}^d , Amer. J. Math. 141 (2019), 857-905.
- [10] F. Bruhat, Distributions sur un groupe localement compact et applications à l'étude des représentations des groupes ρ-adiques, Bull. Soc. Math. France 89 (1961), pp. 43-75.
- [11] A. Calderón, Ergodic theory and translation invariant operators, Proc. Natl. Acad. Sci. USA 59 (1968), pp. 349–353.
- [12] Q. Chu, N. Frantzikinakis, B. Host, Ergodic averages of commuting transformations with distinct degree polynomial iterates, Proc. Lond. Math. Soc. 102 (2011), no. 5, pp. 801–842.
- [13] Y. Do, C. Muscalu, C. Thiele, Variational estimates for paraproducts, Rev. Mat. Iberoam. 28 (2012), no. 3, pp. 857–878.
- [14] J. Duoandikoetxea, J.L. Rubio de Francia, Maximal and singular integral operators via Fourier transform estiamtes, Invent. Math. 84 (1986), no. 3, pp. 541–561.
- [15] G. B. Folland, A course in abstract harmonic analysis, 2nd ed., Textbooks in Mathematics, CRC Press, Boca Raton, FL, 2016.
- [16] N. Frantzikinakis, Some open problems on multiple ergodic averages, Bull. Hellenic Math. Soc. 60 (2016), pp. 41–90.
- [17] N. Frantzikinakis, B. Host, B. Kra, Multiple recurrence and convergence for sequences related to the prime numbers, J. Reine Angew. Math., 611 (2007), 131–144.
- [18] N. Frantzikinakis, B. Kuca, Joint ergodicity for commuting transformations and applications to polynomial sequences, *Invent. Math.* **239** (2025), 621-706.
- [19] N. Frantzikinakis, B. Kra, Polynomial averages converge to the product of integrals, Israel J. Math. 148 (2005), pp. 267–276.
- [20] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, J. Anal. Math. 31 (1977), pp. 204–256.
- [21] L. Grafakos, Classical Fourier Analysis, third edition, Graduate Texts in Mathematics, vol. 249, Springer, New York, 2014.
- [22] W. T. Gowers, A new proof of Szemerédi's theorem, Geom. Funct. Anal., 11 (2001), 465-588.
- [23] B. Green, T. Tao, T. Ziegler, An inverse theorem for the Gowers $U^{s+1}[N]$ -norm, Ann. of Math. (2), 176(2) (2012), 1231-1372.
- [24] B. Host, Ergodic seminorms for commuting transformations and applications, Studia Math. 195 (2009), pp. 31–49.
- [25] B. Host, B. Kra, Non-conventional ergodic averages and nilmanifolds, Ann. of Math. 161 (2005), pp. 397–488.
- [26] B. Host, B. Kra, Convergence of polynomial ergodic averages, Israel J. Math. 149 (2005), pp. 1–19.
- [27] A. Ionescu, S. Wainger, L^p boundedness of discrete singular Radon transforms, J. Amer. Math. Soc. 19 (2005), 357-383.
- [28] H. Iwaniec, E. Kowalski. *Analytic number theory*, Amer. Math. Soc. Colloq. Publ., **53** American Mathematical Society, Providence, RI, 2004.
- [29] D. Kosz, Sharp constants in inequalities admitting the Calderón transference principle, Ergodic Theory Dynam. Systems 44 (2024), pp. 1597–1608.
- [30] D. Kosz, M. Mirek, S. Peluse, J. Wright, The multilinear circle method and a question of Bergelson, arXiv:2411.09478v2.

- [31] B. Krause, M. Mirek, T. Tao, Pointwise ergodic theorems for non-conventional bilinear polynomial averages, Ann. of Math. (2) 195 (2022), 997-1109.
- [32] B. Krause, H. Mousavi, T. Tao, J. Teräväinen, Pointwise convergence of bilinear polynomial averages over the primes, arXiv:2409.10510.
- [33] A. Leibman, Convergence of multiple ergodic averages along polynomials of several variables, Israel J. Math. 146 (2005), pp. 303–315.
- [34] D. Lépingle, La variation d'ordre p des semi-martingales, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete. 36 (1976), no. 4, pp. 295–316.
- [35] K. Matomäki, X. Shao, T. Tao, J. Teräväinen, Higher uniformity of arithmetic functions in short intervals I. All intervals, Forum Math. Pi 11 (2023), Paper No. e29, 97 pp.
- [36] K. Matomäki, X. Shao, Discorrelation between primes in short intervals and polynomial phases, Int. Math. Res. Not., 2021 (2021), 12330–12355.
- [37] N. Mehlhop, W. Słomian, Oscillation and jump inequalities for the polynomial ergodic averages along multi-dimensional subsets of primes, Math. Ann. 388 (2024), no. 3, 2807-2842.
- [38] M. Mirek, Square function estimates for discrete Radon transforms, Anal. PDE 11 (2018), 583-608.
- [39] M. Mirek, E.M. Stein, P. Zorin-Kranich. Jump inequalities via real interpolation, Math. Ann. 376 (2020), no. 1–2, pp. 797–819.
- [40] M. Mirek, E.M. Stein, B. Trojan, ℓ^p -estimates for discrete operators of Radon type: variational estimates, *Invent. Math.* **209** (2017), 665-748.
- [41] M. Mirek, B. Trojan, P. Zorin-Kranich, Variational estimates for averages and truncated singular integrals along the prime numbers, *Trans. Amer. Math. Soc.* **369** (2017), 5403-5423.
- [42] M. Mirek, E.M. Stein, P. Zorin-Kranich, A bootstrapping approach to jump inequalities and their applications, Anal. PDE 13 (2020), no. 2, pp. 527–558.
- [43] R. Nair, On polynomials in primes and J. Bourgain's circle method approach to ergodic theorems, Ergodic Theory Dynam. Systems, 11 (1991), no. 3, 485–499.
- [44] R. Nair, On polynomials in primes and J. Bourgain's circle method approach to ergodic theorems. II., Studia Math., 105 (1993), no. 3, 207–233.
- [45] J. von Neumann, Proof of the quasi-ergodic hypothesis, Proc. Natl. Acad. Sci. USA 18 (1932), pp. 70–82.
- [46] M. S. Osborne, On the Schwartz-Bruhat space and the Paley-Wiener theorem for locally compact abelian groups, J. Functional Analysis 19 (1975), pp. 40–49.
- [47] S. Peluse, Bounds for sets with no polynomial progressions, Forum Math. Pi 8 (2020), e16, 55 pp.
- [48] W. Rudin, Fourier analysis on groups, Interscience Tracts in Pure and Applied Mathematics, No. 12 Interscience Publishers, New York-London 1962.
- [49] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression, Acta Arith. 27 (1975), pp. 199–245.
- [50] T. Tao, Norm convergence of multiple ergodic averages for commuting transformations, Ergodic Theory Dynam. Systems 28 (2008), pp. 657–688.
- [51] T. Tao, The Ionescu-Wainger multiplier theorem and the adeles, Mathematika 67 (2021), 647-677.
- [52] T. Tao, J. Teräväinen, Quantitative bounds for Gowers uniformity of the Möbius and von Mangoldt functions, J. Eur. Math. Soc. (JEMS) 27 (2025), no. 4, 1321-1384.
- [53] J. Teräväinen, Pointwise convergence of ergodic averages with Möbius weight, arXiv:2401.03174v2.
- [54] B. Trojan, Variational estimates for discrete operators modeled on multi-dimensional polynomial subsets of primes, Math. Ann. 374 (2019), 1597-1656.
- [55] M. Walsh, Norm convergence of nilpotent ergodic averages, Ann. of Math. 175 (2012), no. 3, pp. 1667–1688.
- [56] M. Wierdl, Pointwise ergodic theorem along the prime numbers, Israel J. Math., 64 (1988), no. 3, 315–336.
- [57] T. Wooley, T. Ziegler, Multiple recurrence and convergence along the primes, Amer. J. Math, 134 (2012), 1705–1732.
- [58] T. Ziegler, Universal characteristic factors and Furstenberg averages, J. Amer. Math. Soc. 20 (2007), pp. 53–97.

Ministry of Education Key Laboratory of NSLSCS, School of Mathematical Sciences, Nanjing Normal University, Nanjing 210023, People's Republic of China

Email address: wrh@njnu.edu.cn