Annals of Mathematics **195** (2022), 997–1109
https://doi.org/10.4007/annals.2022.195.3.4

Pointwise ergodic theorems for non-conventional bilinear polynomial averages

By Ben Krause, Mariusz Mirek, and Terence Tao

Dedicated to the memory of Jean Bourgain and Elias M. Stein

Abstract

We establish convergence in norm and pointwise almost everywhere for the non-conventional (in the sense of Furstenberg) bilinear polynomial ergodic averages

$$A_N(f,g)(x) := \frac{1}{N}\sum_{n=1}^{N} f(T^n x)g(T^{P(n)}x)$$

as $N \to \infty$, where $T\colon X \to X$ is a measure-preserving transformation of a σ-finite measure space (X,μ), $P(\mathrm{n}) \in \mathbb{Z}[\mathrm{n}]$ is a polynomial of degree $d \geq 2$, and $f \in L^{p_1}(X)$, $g \in L^{p_2}(X)$ for some $p_1, p_2 > 1$ with $\frac{1}{p_1} + \frac{1}{p_2} \leq 1$. We also establish an r-variational inequality for these averages (at lacunary scales) in the optimal range $r > 2$. We are also able to "break duality" by handling some ranges of exponents p_1, p_2 with $\frac{1}{p_1} + \frac{1}{p_2} > 1$, at the cost of increasing r slightly.

This gives an affirmative answer to Problem 11 from Frantzikinakis' open problems survey for the Furstenberg–Weiss averages (with $P(\mathrm{n}) = \mathrm{n}^2$), which is a bilinear variant of Question 9 considered by Bergelson in his survey on Ergodic Ramsey Theory from 1996. This also gives a contribution to the Furstenberg–Bergelson–Leibman conjecture. Our methods combine techniques from harmonic analysis with the recent inverse theorems of Peluse and Prendiville in additive combinatorics. At large scales, the harmonic analysis of the adelic integers $\mathbb{A}_{\mathbb{Z}}$ also plays a role.

Keywords: non-conventional ergodic averages, pointwise convergence, variational estimates, inverse theorems, circle method, adelic integres, Fourier methods

AMS Classification: Primary: 37A30, 37A46, 42A45, 42A50, 42A85, 43A25, 11L03, 11L07, 11L15, 11P55.

Ben Krause was partially supported by the Simons Foundation Analysis and Geometry Research Grant. Mariusz Mirek was partially supported by Department of Mathematics at Rutgers University, and by the National Science Centre in Poland, grant Opus 2018/31/B/ST1/00204. Terence Tao was partially supported by NSF grant DMS-1764034 and by a Simons Investigator Award.

Contents

1. Introduction

1.1. *Non-conventional polynomial ergodic averages.* Define a *measure-preserving system* to be a triple $X = (X, \mu, T)$, where $X = (X, \mu)$ is a σ-finite measure space, and $T\colon X \to X$ is an invertible bimeasurable map that is measure-preserving in the sense that $\mu(T(E)) = \mu(E)$ for all measurable E. In the literature it is common to also require $X = (X, \mu)$ to have finite measure (and often one normalizes (X, μ) to be a probability space), but our main theorem will not require this hypothesis.

Let $\mathbb{Z}[\mathrm{n}]$ denote the space of all formal polynomials $P(\mathrm{n})$ in one indeterminate n with integer coefficients. Such a polynomial $P(\mathrm{n}) \in \mathbb{Z}[\mathrm{n}]$ can of course be identified with a function $P\colon \mathbb{Z} \to \mathbb{Z}$, thus, for instance, n is identified with the identity function $n \mapsto n$ and n^2 is identified with the quadratic function $n \mapsto n^2$. (Later on we will also identify P with maps $P\colon R \to R$ on other commutative rings R, such as the reals $\mathbb{R}$, the p-adic integers $\mathbb{Z}_p$, or the profinite integers $\hat{\mathbb{Z}}$.) Given any polynomials $P_1(\mathrm{n}), \dots, P_k(\mathrm{n}) \in \mathbb{Z}[\mathrm{n}]$, measurable functions $f_1, \dots, f_k \in L^0(X)$ (see Section 2 for a definition of this space), and a real number $N \geq 1$, we can define the non-conventional polynomial ergodic average $A_{N;X}^{P_1(\mathrm{n}),\dots,P_k(\mathrm{n})}(f_1, \dots, f_k) \in L^0(X)$ by the formula

$$A_{N;X}^{P_1(\mathrm{n}),\dots,P_k(\mathrm{n})}(f_1, \dots, f_k)(x) \coloneqq \mathbb{E}_{n\in[N]} f_1(T^{P_1(n)}x) \cdots f_k(T^{P_k(n)}x), \tag{1.1}$$

where $\mathbb{E}_{n\in[N]} f(n) \coloneqq \frac{1}{\lfloor N\rfloor} \sum_{n=1}^{\lfloor N\rfloor} f(n)$. (See Section 2 for a more general definition of this averaging notation.) The terminology "non-conventional" for such multilinear averages was introduced in [35] and is now standard in the ergodic theory literature (see, e.g., [36], [44]). We will usually abbreviate $A_{N;X}^{P_1(\mathrm{n}),\dots,P_k(\mathrm{n})}$ as $A_N^{P_1,\dots,P_k}$ or even A_N when this does not cause confusion. As A_N only depends on the integer part $\lfloor N\rfloor$ of N, one could have restricted N to the positive integers $\mathbb{Z}_+$; however it will be convenient to generalize to real-valued N in order to use certain scaling arguments.

Example 1.2 (Integer shift system). The *integer shift system* $\mathbb{Z} = (\mathbb{Z}, \mu_{\mathbb{Z}}, T_{\mathbb{Z}})$ is the set of integers $\mathbb{Z}$ equipped with counting measure $\mu_{\mathbb{Z}}$ and the shift $T_{\mathbb{Z}}(x) \coloneqq x - 1$. For our purposes, this system will be "universal" for all other measure-preserving systems, in a sense formalized by the Calderón transference principle; see Proposition 3.2(ii). This will be a particularly convenient system to work in due to the extensive Fourier-analytic structure available on the additive group of integers $\mathbb{Z}$, which can be connected, in particular, (in the "major arc" regime) to the corresponding Fourier-analytic structures on other locally compact abelian groups, such as the adelic integers $\mathbb{A}_{\mathbb{Z}}$; see Figure 7. In this system one has

$$A_N^{P_1,\dots,P_k}(f_1, \dots, f_k)(x) = \mathbb{E}_{n\in[N]} f_1(x - P_1(n)) \cdots f_k(x - P_k(n)).$$

Our main results will concern the bilinear averages

$$A_N^{\mathrm{n},P(\mathrm{n})}(f, g)(x) = \mathbb{E}_{n\in[N]} f(T^n x) g(T^{P(n)}x) = \frac{1}{\lfloor N\rfloor} \sum_{n=1}^{\lfloor N\rfloor} f(T^n x) g(T^{P(n)}x)$$

for a given polynomial $P(\mathrm{n}) \in \mathbb{Z}[\mathrm{n}]$, but as motivation we shall also discuss the classical ergodic average

$$A_N^{\mathrm{n}} f(x) = \mathbb{E}_{n\in[N]} f(T^n x) = \frac{1}{\lfloor N\rfloor} \sum_{n=1}^{\lfloor N\rfloor} f(T^n x)$$

and the linear polynomial average

$$A_N^{P(\mathrm{n})} f(x) = \mathbb{E}_{n\in[N]} f(T^{P(n)}x) = \frac{1}{\lfloor N\rfloor}\sum_{n=1}^{\lfloor N\rfloor} f(T^{P(n)}x).$$

A central problem in ergodic theory is to understand convergence in norm and pointwise almost everywhere for the non-conventional polynomial ergodic averages (1.1) as $N\to\infty$. This line of investigations has been initiated in the early 1930s by von Neumann's mean ergodic theorem [70] and Birkhoff's pointwise ergodic theorem [7] (see Theorem 1.7) and led to profound generalizations such as Bourgain's polynomial pointwise ergodic theorem [11], [12], [13] (see Theorem 1.8) and Furstenberg's ergodic proof [33] of Szemerédi's theorem [79]. Furstenberg's proof was also the starting point of the multiple/multilinear ergodic theory (see Theorems 1.15 and 1.16) arising in ergodic Ramsey theory that also motivates this paper. Pointwise convergence is the most natural as well as the most difficult type of convergence to establish. It requires sophisticated tools in analysis, ergodic theory and probability. Especially, the context of pointwise convergence of (1.1) will require an understanding of quantitative forms of pointwise convergence, which we briefly illustrate below.

Given some non-conventional average $A_N(f_1,\dots,f_k)$ of some functions $f_1,\dots,f_k$, with each f_i belonging to some Lebesgue space $L^{p_i}(X)$, one can pose the following questions:

(i) (Norm convergence). Does $A_N(f_1,\dots,f_k)$ converge in $L^p(X)$ norm as $N\to\infty$ for some exponent $p>0$?

(ii) (Almost everywhere convergence). Does $A_N(f_1,\dots,f_k)$ converge pointwise almost everywhere (with respect to μ, of course) as $N\to\infty$?

(iii) (Maximal inequality). Can one bound the $L^p(X)$ norm of the maximal function $\sup_{N\in\mathbb{Z}_+}|A_N(f_1,\dots,f_k)|$, (or equivalently, the $L^p(X;\ell^\infty)$ norm of the sequence of averages $(A_N(f_1,\dots,f_k))_{N\in\mathbb{Z}_+}$) for some $p>0$ in terms of the norms $\|f_i\|_{L^{p_i}(X)}$? More precisely, one is concerned with the following bound:

$$\|\sup_{N\in\mathbb{Z}_+}|A_N(f_1,\dots,f_k)|\|_{L^p(X)} \lesssim_{p_1,\dots,p_k,p} \|f_1\|_{L^{p_1}(X)}\cdots\|f_k\|_{L^{p_k}(X)}. \tag{1.3}$$

(See Section 2 for the asymptotic notation used in this paper.)

(iv) (Variational inequality). Can one bound the $L^p(X)$ norm of the r-variational norm $\|(A_N(f_1,\dots,f_k))_{N\in\mathbb{Z}_+}\|_{\boldsymbol{V}^r}$, (or equivalently, the $L^p(X;\boldsymbol{V}^r)$ norm of the sequence of averages $(A_N(f_1,\dots,f_k))_{N\in\mathbb{Z}_+}$) for some $p>0$ and some $1\le r<\infty$ in terms of the norms $\|f_i\|_{L^{p_i}(X)}$? More precisely, one is concerned with the following bound:

$$\big\|\|(A_N(f_1,\dots,f_k))_{N\in\mathbb{Z}_+}\|_{\boldsymbol{V}^r}\big\|_{L^p(X)} \lesssim_{p_1,\dots,p_k,p,r} \|f_1\|_{L^{p_1}(X)}\cdots\|f_k\|_{L^{p_k}(X)}. \tag{1.4}$$

The r-variational norm is defined by

$$\|(A_N(f_1,\dots,f_k))_{N\in\mathbb{Z}_+}\|_{\boldsymbol{V}^r} := \sup_{N\in\mathbb{Z}_+} |A_N(f_1,\dots,f_k)| + \|(A_N(f_1,\dots,f_k))_{N\in\mathbb{Z}_+}\|_{V^r},$$

where $\|(A_N(f_1,\dots,f_k))_{N\in\mathbb{Z}_+}\|_{V^r}$ is given by the following expression

$$\sup_{J\in\mathbb{Z}_+} \sup_{\substack{N_0\le\cdots\le N_J\\ N_j\in\mathbb{Z}_+}} \Big(\sum_{j=0}^{J-1} |A_{N_{j+1}}(f_1,\dots,f_k) - A_{N_j}(f_1,\dots,f_k)|^r\Big)^{1/r}; \tag{1.5}$$

here the supremum is taken over all finite increasing sequences in $\mathbb{Z}_+$. (See Section 2 for a more general definition of the variational norm $\boldsymbol{V}^r$ and its properties.)

These questions are all related to each other. For instance, if variational inequality (1.4) holds, then one automatically has a maximal inequality (1.3). Moreover, (1.4) immediately ensures that the quantity in (1.5) is finite almost everywhere, which in turn implies almost everywhere convergence of the sequence $(A_N(f_1,\dots,f_k))_{N\in\mathbb{Z}_+}$ as $N\to\infty$. Norm convergence then also follows (for $p<\infty$) by (1.3) and the dominated convergence theorem. This variational norm approach to ergodic theorems was advocated, in particular, by Bourgain [12], and it is very useful in pointwise convergence problems with arithmetic features.

We say that a tuple $(p_1,\dots,p_k,p)$ of exponents is *Hölder* if $\frac{1}{p}=\frac{1}{p_1}+\cdots+\frac{1}{p_k}$ and *Banach* if $p_1,\dots,p_k,p\ge 1$. If a tuple $(p_1,\dots,p_k,p)$ is both Hölder and Banach, then from Hölder's inequality and the triangle inequality in the Banach space $L^p(X)$ one has

$$\|A_N(f_1,\dots,f_k)\|_{L^p(X)} \le \|f_1\|_{L^{p_1}(X)}\cdots\|f_k\|_{L^{p_k}(X)} \tag{1.6}$$

regardless of the choice of polynomials $P_1(\mathrm{n}),\dots,P_k(\mathrm{n})$. Thus it is natural to restrict attention to the case of exponents that are both Hölder and Banach. The Hölder hypothesis is particularly essential for ergodic theory applications as it is needed in order to apply the Calderón transference principle; see Proposition 3.2(ii). However, we will be able to "break duality" in our main result by allowing certain non-Banach exponents $p<1$ while still maintaining the Hölder property; see Section 11. On the integer shift model $\mathbb{Z}$, the estimates become trivial (and of little use) in the super-Hölder regime $\frac{1}{p}<\frac{1}{p_1}+\cdots+\frac{1}{p_k}$, and false in the opposite sub-Hölder regime $\frac{1}{p}>\frac{1}{p_1}+\cdots+\frac{1}{p_k}$; see Remark 3.11.

It is technically convenient to sparsify the set of scales N that one is ranging over to define a maximal or variational function. For instance, one could replace the positive integers $\mathbb{Z}_+$ by the dyadic integers

$$2^{\mathbb{N}} := \{2^k : k\in\mathbb{N}\}.$$

More generally, we can work with sets $\mathbb{D} = \{N_1, N_2, \dots\}$ of positive reals $1 \leq N_1 < N_2 < \cdots$ that are λ-*lacunary* for some $\lambda > 1$, in the sense that

$$N_{j+1}/N_j > \lambda$$

for all $j \in \mathbb{Z}_+$; one defines λ-lacunarity for finite sequences $\{N_1, \dots, N_k\}$ of positive reals $1 \leq N_1 < \cdots < N_k$ in a similar fashion. Variational estimates on such lacunary sets are sometimes referred to as "long variation estimates" in the literature; they are somewhat weaker than full variation estimates but are often still sufficient for applications such as demonstrating almost everywhere convergence.

We will only concern ourselves in this paper with the existence of a limit of an ergodic average, and not attempt to compute what the limiting average actually is. The nature of this limiting average is now fairly well understood (at least when $f_1, \dots, f_k \in L^\infty(X)$ and X has finite measure) thanks to the theory of characteristic factors and the equidistribution theory of nilmanifolds; see, for instance, [3], [4], [30] for further discussion. In particular, for a description of the limit in the case when the polynomials all have distinct degrees, which is of course the case of primary interest here, we refer to [19]. We also remark that the limit in this case is determined entirely by the projection of the functions to the rational factor (the factor spanned by periodic functions), which is the ergodic theory analogue of the "major arc" component of the functions. These results are also related to recurrence and Roth and Szemerédi type theorems (see, e.g., [33], [35], [34], [5], [79]), which also motivate this paper, but we will not discuss these topics further here.

1.2. *Linear averages.* We now recall the standard ergodic theorems for the classical ergodic averages A_N^{n}:

THEOREM 1.7 (Classical ergodic averages). *Let $X = (X, \mu, T)$ be a measure-preserving system, and let $f \in L^p(X)$ for some $1 \leq p \leq \infty$.*

(i) (Mean ergodic theorem). *If $1 < p < \infty$, then $A_N^{\mathrm{n}} f$ converges in $L^p(X)$ norm.*

(ii) (Pointwise ergodic theorem). *If $1 \leq p < \infty$, then $A_N^{\mathrm{n}} f$ converges pointwise almost everywhere.*

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

$$\|(A_N^{\mathrm{n}} f)_{N \in \mathbb{Z}_+}\|_{L^p(X; \ell^\infty)} \lesssim_p \|f\|_{L^p(X)}.$$

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

$$\|(A_N^{\mathrm{n}} f)_{N \in \mathbb{Z}_+}\|_{L^p(X; \boldsymbol{V}^r)} \lesssim_{p,r} \|f\|_{L^p(X)}.$$

Proof. Parts (i)–(iii) are standard, particularly in the case when X has finite measure, and are due to von Neumann [70], Birkhoff [7], and Hopf [42]; the maximal inequality (for σ-finite X) can also be established by transference

to the integer shift case $(\mathbb{Z}, \mu_{\mathbb{Z}}, T_{\mathbb{Z}})$ and then applying the Hardy–Littlewood maximal inequality. (This also gives a weak-type endpoint for (iii).) The variational estimate was established by Bourgain [12, Cor. 3.26] in the $p = 2$ case, and the general case was established in [49]; this estimate can then be used to recover the mean and pointwise ergodic theorems in the σ-finite case as mentioned previously. □

We have (slightly weaker) analogues of these results for other linear polynomial averages:

THEOREM 1.8 (Linear polynomial averages). *Let $X = (X, \mu, T)$ be a measure-preserving system, let $P(\mathrm{n}) \in \mathbb{Z}[\mathrm{n}]$, and let $f \in L^p(X)$ for some $1 \le p \le \infty$.*

(i) (Mean ergodic theorem). *If $1 < p < \infty$, then $A_N^{P(\mathrm{n})} f$ converges in $L^p(X)$ norm.*

(ii) (Pointwise ergodic theorem). *If $1 < p < \infty$, then $A_N^{P(\mathrm{n})} f$ converges pointwise almost everywhere.*

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

$$\|(A_N^{P(\mathrm{n})} f)_{N \in \mathbb{Z}_+}\|_{L^p(X;\ell^\infty)} \lesssim_{p,P} \|f\|_{L^p(X)}. \tag{1.9}$$

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

$$\|(A_N^{P(\mathrm{n})} f)_{N \in \mathbb{Z}_+}\|_{L^p(X;\boldsymbol{V}^r)} \lesssim_{p,r,P} \|f\|_{L^p(X)}. \tag{1.10}$$

Proof. Part (i) follows for $p = 2$ by a routine application of the spectral theorem (or one can invoke Theorem 1.15 below), and the other values of p then follow from a density argument. Parts (ii), (iii) were established by Bourgain [12, Th. 1] (see also [11], [13]). Part (iv) was established in the $p = 2$ case by the first author in [54, Prop. 1.5] by adapting the methods of Bourgain, and in full generality by the second author and his collaborators in [66]; see also [68]. In [54, §8] it is also shown that (1.10) fails at the endpoint $p = r = 2$. For $p = 1$, in contrast to Theorem 1.7(ii), pointwise convergence in Theorem 1.8(ii) fails for any monomial $P(\mathrm{n}) = \mathrm{n}^d$ of degree $d \ge 2$, as was shown in [17], [56]. □

Theorem 1.8 is proven via the circle method. The implementation of this method can be summarized in the following two sentences:

(i) Plancherel's theorem and Weyl sum estimates are used to control the contribution of minor arcs.

(ii) Multifrequency harmonic analysis is used to control the contribution of major arcs.

We now briefly sketch some more details of Bourgain's proof for maximal inequality (1.9). The key estimate to establish is (1.9) when $p = 2$ and (X, μ, T) is the integer shift system, where N is restricted to a finite lacunary set $\mathbb{I}$, and

with f assumed to be in the Schwartz–Bruhat space $\boldsymbol{S}(\mathbb{Z}) \subset \ell^1(\mathbb{Z})$ to avoid technicalities; see Section 4 for a definition of this space. In this setting we have the convenient Fourier representation

$$\mathcal{F}_{\mathbb{Z}} A_N^{P(\mathrm{n})} f(\xi) = \varphi_{N,\mathbb{Z}}(\xi) \mathcal{F}_{\mathbb{Z}} f(\xi)$$

for any $\xi \in \mathbb{T}$, where using the averaging notation (2.2) the symbol $\varphi_{N,\mathbb{Z}}(\xi)$ is given by

$$\varphi_{N,\mathbb{Z}}(\xi) := \mathbb{E}_{n \in [N]} e(P(n)\xi), \tag{1.11}$$

where $e(\theta) := e^{2\pi i \theta}$ and the Fourier transform $\mathcal{F}_{\mathbb{Z}} f$ are defined in Section 4. Standard Weyl sum estimates (see [48, Lemma 20.3, p. 462]) reveal that for some small $\delta, \varepsilon > 0$, one has

$$|\varphi_{N,\mathbb{Z}}(\xi)| \lesssim_P N^{-\delta}, \tag{1.12}$$

unless ξ is in a *major arc*, which roughly speaking means that ξ is close to $\frac{a}{q} \bmod 1$ for some $a \in \mathbb{Z}$ and some small positive integer $1 \leq q \leq N^{\varepsilon}$. One can then use (1.12) and Plancherel's theorem to dispose of the *minor arc* case when ξ is not in a major arc, and then after a dyadic decomposition the main task is to establish an estimate roughly of the shape

$$\|(A_N^{P(\mathrm{n})} f)_{N \in \mathbb{I}}\|_{\ell^2(\mathbb{Z};\ell^\infty)} \lesssim_{r,P,\lambda} 2^{-cl} \|f\|_{\ell^2(\mathbb{Z})}$$

for all $l \in \mathbb{N}$, $\lambda > 1$ and some constant $c = c_{r,P} > 0$, where $\mathbb{I} \subset [1, +\infty)$ is an arbitrary finite λ-lacunary set and the Fourier transform of f is restricted to the set of "l-major arc" frequencies ξ of the form $\xi = \frac{a}{q} + O(2^{-10l}) \bmod 1$ (say) for some $q \sim 2^l$. (Informally, this is morally equivalent by the uncertainty principle to f being a linear combination of functions that are approximately constant on arithmetic progressions of spacing q for various $q \sim 2^l$ and diameter $\sim 2^{10l}$; see Remark 5.20.) In fact, at a given (large) scale N one can restrict to even narrower major arcs, of width $O(2^{dl}/N^d)$ say. A finer analysis of the symbol (1.11) reveals that for a major arc frequency $\xi = \frac{a}{q} + \theta \bmod 1$,

$$\varphi_{N,\mathbb{Z}}\left(\frac{a}{q} + \theta \bmod 1\right)$$

has an approximate factorization

$$\varphi_{\hat{\mathbb{Z}}}\left(\frac{a}{q} \bmod 1\right) \varphi_{N,\mathbb{R}}(\theta), \tag{1.13}$$

where the "arithmetic symbol" $\varphi_{\hat{\mathbb{Z}}} \colon \mathbb{Q}/\mathbb{Z} \to \mathbb{C}$ is defined by

$$\varphi_{\hat{\mathbb{Z}}}\left(\frac{a}{q} \bmod 1\right) := \mathbb{E}_{n \in \mathbb{Z}/q\mathbb{Z}} e\left(\frac{aP(n)}{q}\right) \tag{1.14}$$

and the "continuous symbol" $\varphi_{N,\mathbb{R}}\colon \mathbb{R} \to \mathbb{C}$ is defined by

$$\varphi_{N,\mathbb{R}}(\theta) := \frac{1}{N}\int_0^N e(\theta P(t))\, dt.$$

The influence of the arithmetic symbol $\varphi_{\hat{\mathbb{Z}}}$ (which does not depend on N) can be easily factored out in the $p = 2$ case by Plancherel's theorem, and the task then readily reduces to that of establishing a multifrequency maximal inequality (see [13, Lemma 4.1]). This result in turn is ultimately derived from a variational inequality for averages of vector-valued L^2 functions (see [13, Lemma 3.30]), in the spirit of Lépingle's inequality.

1.3. *Bilinear averages.* Now we turn to multilinear averages. For the norm convergence problem in the case of finite measure and Banach exponents the situation is well understood, thanks to the following result of Host–Kra and Leibman:

THEOREM 1.15 (Multilinear mean ergodic theorem). *Let (X, μ, T) be a measure-preserving system of finite measure, let $P_1(\mathrm{n}), \dots, P_k(\mathrm{n}) \in \mathbb{Z}[\mathrm{n}]$, and let $f_i \in L^{p_i}(X)$ for all $i = 1, \dots, k$ and some exponents $1 \le p_i \le \infty$ with $\frac{1}{p_1} + \cdots + \frac{1}{p_k} \le 1$. Then the averages $A_N^{P_1,\dots,P_k}(f_1, \dots, f_k)$ converge in $L^p(X)$ norm for any $0 < p < \infty$ with $\frac{1}{p_1} + \cdots + \frac{1}{p_k} < \frac{1}{p}$.*

Proof. The case $p_1 = \cdots = p_k = \infty$, $p = 2$ is established in [43], [57] (see also [82] and [1] for quite different proofs and generalizations); one can then extend to other $0 < p < \infty$ by Hölder's inequality, and the case of general $p_1, \dots, p_k$ then follows by a standard limiting argument using (1.6). □

There is a long history of prior partial results (e.g., [44], [83], [2], [5], [31], [36]) towards Theorem 1.15, as well as generalizations to actions of other nilpotent groups than $\mathbb{Z}$ (i.e., averages involving multiple measure-preserving transformations $T_1, \dots, T_k$ that generate a nilpotent group); we refer the reader to [3], [4], [30] for surveys. In several cases it is possible to "break duality" by permitting $\frac{1}{p_1} + \cdots + \frac{1}{p_k}$ to exceed 1; see Section 11 below.

For pointwise convergence and for two linear polynomials, one also has the following results:

THEOREM 1.16 (Two linear polynomials). *Let (X, μ, T) be a measure-preserving system with finite measure, let $P_1(\mathrm{n}), P_2(\mathrm{n}) \in \mathbb{Z}[\mathrm{n}]$ have degree 1 with distinct leading coefficients, and let $1 < p_1, p_2 \le \infty$ be such that $\frac{1}{p_1} + \frac{1}{p_2} < \frac{3}{2}$. Then for $f \in L^{p_1}(X), g \in L^{p_2}(X)$, the averages $A_N^{P_1,P_2}(f, g)$ converge pointwise almost everywhere.*

Proof. For the case $p_1 = p_2 = \infty$, see Bourgain [14]; an alternate proof was also given by Demeter [21]. To extend to the remaining cases of p_1, p_2

one applies a bilinear maximal inequality of Lacey [55] and a standard limiting argument. □

We now at last come to the main result of our paper, which concerns an opposing case to Theorem 1.16 in which one has one linear polynomial and one strictly nonlinear polynomial.

THEOREM 1.17 (Main theorem). *Let (X,μ,T) be a measure-preserving system, let $P(\mathrm{n})\in\mathbb{Z}[\mathrm{n}]$ have degree $d\geq 2$, and let $f\in L^{p_1}(X), g\in L^{p_2}(X)$ for some $1<p_1,p_2<\infty$ with $\frac{1}{p_1}+\frac{1}{p_2}=\frac{1}{p}\leq 1$.*

(i) (*Mean ergodic theorem*) *The averages $A_N^{\mathrm{n},P(\mathrm{n})}(f,g)$ converge in $L^p(X)$ norm.*

(ii) (*Pointwise ergodic theorem*) *The averages $A_N^{\mathrm{n},P(\mathrm{n})}(f,g)$ converge pointwise almost everywhere.*

(iii) (*Maximal ergodic theorem*) *One has*

$$\|(A_N^{\mathrm{n},P(\mathrm{n})}(f,g))_{N\in\mathbb{Z}_+}\|_{L^p(X;\ell^\infty)}\lesssim_{p_1,p_2,P}\|f\|_{L^{p_1}(X)}\|g\|_{L^{p_2}(X)}.$$

(iv) (*Long variational ergodic theorem*) *If $r>2$ and $\lambda>1$, one has*

$$\|(A_N^{\mathrm{n},P(\mathrm{n})}(f,g))_{N\in\mathbb{D}}\|_{L^p(X;\boldsymbol{V}^r)}\lesssim_{p_1,p_2,r,P,\lambda}\|f\|_{L^{p_1}(X)}\|g\|_{L^{p_2}(X)} \tag{1.18}$$

whenever $\mathbb{D}\subset[1,+\infty)$ is λ-lacunary.

We now give some remarks about this theorem.

(i) Theorem 1.17(i) already follows from Theorem 1.15 when (X,μ) has finite measure; in fact for this particular average, the results are essentially already contained in [36]. However, it appears to be new in the σ-finite setting, and the proof method is completely different from methods used to establish Theorem 1.15.

(ii) Theorem 1.17(ii) is completely new for general measure-preserving systems,[1] even when $f,g\in L^\infty(X)$ and X has finite measure. In particular, Theorem 1.17(ii) when specialized to the case $P(\mathrm{n})=\mathrm{n}^2$ answers the second part of [30, Prob. 11] for the Furstenberg–Weiss averages [36] (see also [35]), which is a bilinear variant of the problem considered

[1] This result (and also part (iii)) was claimed in [29]. However, there appear to be several gaps in the arguments. Firstly, in [29, pp. 23] it is claimed without giving details that the Caldéron transference principle can be applied for the super-Hölder exponent triplet $\ell^2\times\ell^2\to\ell^2$, but if one carefully works through the arguments provided in [29, pp. 10–11] for these exponents, one loses a factor of $N^{1/2}$ in the estimates (as h now needs to be controlled in ℓ^2 norm rather than ℓ^∞ norm) and thus cannot pass to the limit $N\to\infty$. Secondly, in [29, pp. 26], bilinear maximal estimates are obtained for the super-Hölder exponent triplets $\ell^1\times\ell^r\to\ell^r$ and $\ell^r\times\ell^1\to\ell^r$, but the assertion in that paper that bilinear interpolation then gives Hölder exponent estimates such as $\ell^1\times\ell^\infty\to\ell^1$ or $\ell^r\times\ell^{r'}\to\ell^1$ is false.

by Bergelson [3, Question 9, pp. 52]; see also [4, §6, pp. 838]. Theorem 1.17 is also a contribution towards establishing the Furstenberg–Bergelson–Leibman conjecture [6, §5.5, p. 468], which asserts the following. Given integers $d, k, m, N \in \mathbb{Z}_+$, let $T_1, \dots, T_d : X \to X$ be a family of invertible measure-preserving transformations of a probability measure space (X, μ) that generates a nilpotent group of step m. Assume that $P_{1,1}, \dots, P_{i,j}, \dots, P_{d,k} \in \mathbb{Z}[\mathrm{n}]$. Then for any $f_1, \dots, f_k \in L^\infty(X)$, the non-conventional multiple polynomial averages

$$\mathbb{E}_{n\in[N]} \prod_{j=1}^{k} f_j(T_1^{P_{1,j}(n)} \dots T_d^{P_{d,j}(n)} x)$$

converge pointwise for μ-almost every $x \in X$ as $N \to \infty$. This conjecture is a widely open problem in ergodic theory that was promoted in person by Furstenberg (see [1, p. 6662] and [53]) before being published in [6]. Bergelson–Leibman [6] showed that convergence may fail if the transformations $T_1, \dots, T_d$ generate a solvable group. Our main theorem solves this conjecture in the case $d = 1$, $k = 2$ with $P_{1,1}(\mathrm{n}) = \mathrm{n}$ and $P_{1,2}(\mathrm{n}) = P(\mathrm{n}) \in \mathbb{Z}[\mathrm{n}]$ with $\deg P \geq 2$. Pointwise convergence for non-conventional polynomial averages has previously been established for some special measure-preserving systems, such as exact endomorphisms and K-automorphisms [24] and nilsystems [58].

(iii) Our methods of proofs break down in the linear case $d = 1$ (as the minor arc contributions are no longer negligible), and so we are unable to give an alternate proof of Theorem 1.16.

(iv) For $p > 1$ (i.e., above the line of duality), Theorem 1.17(iii) follows easily from past results. Indeed, from several applications of Hölder's inequality one has

$$\begin{aligned}\|(A_N^{\mathrm{n},P(\mathrm{n})}(f,g))_{N\in\mathbb{Z}_+}\|_{L^p(X;\ell^\infty)} &\\ \leq \|(A_N^{\mathrm{n}}(|f|^{p_0})|)_{N\in\mathbb{Z}_+}\|^{1/p_0}_{L^{p_1/p_0}(X;\ell^\infty)} &\|(A_N^{P(\mathrm{n})}(|g|^{p_0'})|)_{N\in\mathbb{Z}_+}\|^{1/p_0'}_{L^{p_2/p_0'}(X;\ell^\infty)}\end{aligned}$$

for any $1 < p_0 < \infty$. In the $p > 1$ case one can select p_0 so that $p_0 < p_1$ and $p_0' < p_2$, and the claim now follows from Theorem 1.8(iii). However, the $p = 1$ case (i.e., on the line of duality) is new, even when $p_1 = p_2 = 2$. Also, the simple argument given above does not seem to easily adapt to give the $p > 1$ cases of the other components (i), (ii), (iv) of the theorem, although it does permit one to reduce those cases of (i), (ii) to the case in which $f, g \in L^\infty(X)$ by the usual limiting argument. A continuous analogue of Theorem 1.17(iii) was previously established in [61] (see also [60], [37]).

(v) Theorem 1.17(iv) is the key result in the theorem, and it easily implies the other parts of the theorem, as we shall show in Section 3. The condition $r > 2$ is necessary, as no variational estimate is possible for $r \leq 2$; see Corollary 12.4. The situation can be contrasted with that in [25], in which a certain bilinear paraproduct was shown to enjoy r-variation estimates for some values of $r < 2$.

(vi) A modification of our arguments (taking particular advantage of linear L^p improving estimates) is able to "break duality" and establish some cases of Theorem 1.17 in the non-Banach regime $p < 1$, with the range of exponents being particularly strong in the case of norm convergence on spaces of finite measure; see Section 11. A similar "breaking duality" phenomenon occurred in [22]; also, in [61, Th. 2] a continuous analogue of part (iii) of the theorem was established that "broke duality" by allowing p to lie in the range $p > \frac{d-1}{d}$, which is best possible up to the endpoint; see [61, §3].

(vii) The requirement that X be σ-finite can be dropped by observing that $f \in L^{p_1}(X), g \in L^{p_2}(X)$ have σ-finite supports (since $p_1, p_2 < \infty$), and hence the invariant set $\bigcup_{n\in\mathbb{Z}} T^n(\mathrm{supp}(f) \cup \mathrm{supp}(g))$ is also σ-finite. Since the averages $A_N^{\mathrm{n},P(\mathrm{n})}(f,g)$ are all supported on this invariant σ-finite set, one can restrict to the σ-finite case without loss of generality.

1.4. *Overview of proof.* We now give an overview of the proof of Theorem 1.17. The arguments follow the basic framework of the arguments used to establish the linear results in Theorem 1.8, but with several new difficulties arising that require substantial new ideas to overcome. Most notably,

(a) Plancherel's theorem and Weyl sum estimates ([48, Lemma 20.3, p. 462]) are no longer sufficient by themselves to control the contribution of the minor arcs, thus defeating a "naive" implementation of the circle method.

(b) The bilinear analogue

$$m_{\hat{\mathbb{Z}}}\left(\frac{a_1}{q} \bmod 1, \frac{a_2}{q} \bmod 1\right) := \mathbb{E}_{n\in\mathbb{Z}/q\mathbb{Z}} e\left(\frac{a_1 n + a_2 P(n)}{q}\right) \tag{1.19}$$

of the arithmetic symbol $\varphi_{\hat{\mathbb{Z}}}$ defined in (1.14) cannot be factorized as a tensor product of a function of $\frac{a_1}{q}$ mod 1 and a function of $\frac{a_2}{q}$ mod 1. As a consequence, symbol (1.19), despite being independent of N, cannot be disposed of purely by linear tools such as Plancherel's theorem (even in the model case $p_1 = p_2 = 2$) due to a bilinear nature of the problem and must be treated in tandem with the continuous features.

Our resolution to these problems can be summarized by the following two sentences:

(i) Additive combinatorics (and, more specifically, Peluse–Prendiville theory), as well as Hahn–Banach separation theorem, Ionescu–Wainger multiplier theory, and the $\ell^p(\mathbb{Z})$ improving theory of Han–Kovač–Lacey–Madrid–Yang on the integers $\mathbb{Z}$, are used to control the contribution of minor arcs. This is a bilinear theory of the minor arcs.

(ii) Adelic harmonic analysis (which combines the continuous harmonic analysis of the reals $\mathbb{R}$ with the arithmetic harmonic analysis of the profinite integers $\hat{\mathbb{Z}}$), as well as Ionescu–Wainger multiplier theory, two-parameter Rademacher–Menschov argument, shifted square function estimates, and the $L^p(\hat{\mathbb{Z}})$ improving theory on the profinite integers $\hat{\mathbb{Z}}$, are used to control the contribution of major arcs.

We now discuss the strategy in more detail.

1.4.1. *Standard reductions.* Following Bourgain [11], it suffices to establish the variational estimate (3.3) on a finite λ-dyadic set $\mathbb{D}$ of scales, and by using the Calderón transference principle we can work with the integer shift system $\mathbb{Z}$. For technical reasons, it is also convenient to remove the lower half $n \leq N/2$ of the averaging operator (1.1) and only retain the upper half $n > N/2$, but we ignore this step for the sake of discussion. These standard reductions are reviewed in Section 3. We will need to establish the variational estimate for all choices of (p_1, p_2), but the most important case is when $p_1 = p_2 = 2$ (and hence $p = 1$), where it is easiest to establish a certain exponential decay that can then be propagated to all other choices of exponents (p_1, p_2) by interpolation. For the sake of discussion, we therefore restrict attention to the $p_1 = p_2 = 2$ case.

1.4.2. *Minor arcs estimates.* Again following Bourgain, we would now like to restrict the functions f, g to major arcs in Fourier space. In the linear setting this could be accomplished relatively easily using Plancherel's theorem and decay estimates (1.12) for the symbol (1.11) on minor arcs. However, in the bilinear setting Plancherel's theorem and the classical Weyl estimate [48, Lemma 20.3, p. 462] are insufficient to obtain satisfactory control on the minor arc contribution. Instead we use a deep recent inverse theorem of Peluse and Prendiville [74] and Peluse [73] from the additive combinatorics literature (see Theorem 6.3), which asserts that for every $0 < \delta \leq 1$ and bounded functions $f, g : [-O(N^d), O(N^d)] \to \mathbb{C}$ with $\|f\|_{\ell^\infty}, \|g\|_{\ell^\infty} \leq 1$, if $\|A_N(f,g)\|_{\ell^1} \geq \delta N^d$, then f must weakly correlate with the indicator function of a progression $P = \{qm \in \mathbb{Z}_+ : m \in [N']\}$ with $q \lesssim \delta^{-O(1)}$ and $\delta^{O(1)} N \lesssim N' \leq N$, in the sense that $\|f * \mathbb{1}_{-P}\|_{\ell^1} \gtrsim \delta^{O(1)} N' N^d$ provided that $N \gtrsim \delta^{-O(1)}$. In other words, it says that the function f has a major arc structure at scale N, which is precisely stated using Fourier-transform language in Proposition 6.6. However, for our application, we need to replace the ℓ^∞ control with (suitably normalized) ℓ^2 control. To do this we shall use the Hahn–Banach theorem to interpret this

inverse theorem as a structural description of certain *dual functions* associated to the averaging operator A_N; see Corollary 6.10. We also need to utilize the multiplier theory of Ionescu and Wainger [47] to maintain the separation of major and minor arcs during this process; see Proposition 6.15.

Then combine the latter with recent linear L^p-improving estimates on $\mathbb{Z}$ by Han–Kovač–Lacey–Madrid–Yang [41] (see also Dasu–Demeter–Langowski [20]) to relax the hypotheses to ℓ^2; see Corollary 6.22. The final conclusion of this analysis is the single-scale minor arc estimate in Theorem 5.12, which roughly speaking (with the notation from (2.1)) asserts that

$$\|A_N(f,g)\|_{\ell^1} \lesssim (2^{-O(l)} + \langle \operatorname{Log} N\rangle^{-O(1)})\|f\|_{\ell^2}\|g\|_{\ell^2} \tag{1.20}$$

unless the Fourier transform of f and g are supported on major arcs of width respectively $O(2^l N^{-1})$ and $O(2^{dl}N^{-d})$. (The disparity is due to the different degrees in the polynomials $\mathrm{n}, P(\mathrm{n})$.) Inequality (1.20) can be thought of as a bilinear variant of inequality (1.12), which was derived from classical Weyl's inequality [48, Lemma 20.3, p. 462]. This bilinear inequality (1.20) is a very useful result that we will apply repeatedly in our arguments.

1.4.3. *Major arcs estimates: a first glimpse.* One can now restrict attention to major arcs, in which f has Fourier support supported at combinations $\alpha+\theta$ mod 1 of "arithmetic frequencies" $\alpha\in\mathbb{Q}/\mathbb{Z}$ and "continuous frequencies" $\theta\in\mathbb{R}$. The "height" of the arithmetic frequency α will be bounded by some threshold 2^{l_1}, and the magnitude $|\theta|$ of the continuous frequency will similarly be bounded by some threshold 2^{k_1} for some large negative k_1. With some additional effort, g can similarly be restricted to major arc frequencies that are the combination of an arithmetic frequency of height at most 2^{l_2} and a continuous frequency of magnitude at most 2^{k_2}. Naively, the height of an arithmetic frequency $\alpha = \frac{a}{q}$ mod 1 with $(a,q)=1$ might be defined to equal q (or $\inf\{2^l : q\le 2^l\}$, if one wishes to view height as a dyadic integer). However, for technical reasons, it is often more convenient to replace this naive notion of height with a more complicated variant of height implicitly introduced by Ionescu and Wainger [47] that enjoys better multiplier theory (the losses incurred here are only polynomial in l rather than exponential); see Appendix A. In order to decouple the continuous aspects of the analysis from the arithmetic aspects, it turns out to be convenient to embed the integers $\mathbb{Z}$ into the adelic integers[2] $\mathbb{A}_{\mathbb{Z}} := \mathbb{R}\times\hat{\mathbb{Z}} = \mathbb{R}\times\prod_p \mathbb{Z}_p$; this embedding $\iota\colon \mathbb{Z}\to\mathbb{A}_{\mathbb{Z}}$ is the Fourier

[2]One could also work with various projections $\mathbb{R}\times\mathbb{Z}/Q\mathbb{Z}$ of the adelic integers, which amounts to requiring a common denominator Q to the arithmetic frequencies being used; but the adelic formalism is cleaner in that it automatically handles uniformity in the Q parameter. Also we believe it lends some conceptual clarity to the strategy of separating the continuous and arithmetic aspects of the analysis.

adjoint of the addition map $\pi\colon \mathbb{R}\times\mathbb{Q}/\mathbb{Z}\to\mathbb{T}$ defined by $\pi(\theta,\alpha) := \alpha+\theta$ that was implicitly used to define major arcs. The advantage of working in the adelic framework is that several key linear and bilinear Fourier symbols on the integers, when transferred to the adelic integers, can be treated in a fairly unified way and can be cleanly decomposed or approximated into simpler symbols that exhibit a useful tensor product structure, so that the continuous and arithmetic aspects of the symbols involved become almost completely decoupled; see also Figures 1 and 2.

1.4.4. *Major arcs estimates*: *paraproduct-type decomposition.* The objective is now to obtain, for a given choice of height scales l_1, l_2, variational bounds on the average $A_N(f,g)$ under the assumption that f, g have Fourier supports associated to major arcs of heights $2^{l_1}, 2^{l_2}$ respectively, with the bounds enjoying exponential decay in the parameter $l := \max(l_1, l_2)$. At a given scale N, one can use the Ionescu–Wainger multiplier theory to restrict the Fourier transform of f to major arcs of width about $2^l N^{-1}$, and similarly restrict the Fourier transform of g to major arcs of width about $2^{dl}N^{-d}$. (As before the disparity is due to the different degrees in the polynomials $\mathrm{n}, P(\mathrm{n})$.) For any given scale N, Theorem 5.12 gives the desired exponential gain in l; the problem is how to sum in N. To overcome this difficulty we perform a certain paraproduct decomposition (5.27) and (5.28) centered around a finite number of (arithmetic) frequencies. This contrasts sharply with the classical theory of paraproducts that are centered at the frequency origin. Here, again an indispensable role is played by the Ionescu–Wainger projections (5.15) and (5.16), which will allow us to control "low-low," "low-high," "high-low," and "high-high" paraproducts by employing the methods from continuous harmonic analysis.

1.4.5. *Major arcs estimates*: "*low-low*" *case and* "*small scales*". For the sake of exposition, let us initially focus on the "low-low" case when one can restrict the width of the major arcs further to $2^{-u}N^{-1}$ and $2^{-du}N^{-d}$ where u is moderately large (about $2^{\rho l}$ for some small constant ρ). The argument then splits into the treatment of "small scales" $2^u < N < 2^{2^u}$ and "large scales" $N > 2^{2^u}$. (The contribution of extremely small scales $N \le 2^u$ can be easily discarded, thanks to the exponential decay factors present in the single scale estimates). For small scales, in the linear theory we used the Rademacher–Menshov type inequality [66], which was quite efficient. Here, due to the bilinear nature of the problem the situation is much more complicated. We begin with performing some Fourier-analytic approximations at the adelic integer level, analogous to (1.13), to replace averages such as $A_N(f,g)$ with an expressions of the form $B(f_N, g_N)$, where the bilinear operator B is now independent of N. This is the key idea of the major arcs analysis, which is encapsulated in the model estimate (7.24) of Theorem 7.23. The same idea is also exploited in the "large scales" to establish estimate (7.25) of Theorem 7.23. After these approximations, we use

a two-parameter Rademacher–Menshov argument and Khinchine's inequality to reduce the variational estimates to a single scale estimates; such arguments lose factors that are essentially logarithmic in the number of scales, which in the small scale regime gives a loss of $u^{O(1)}$, but this is acceptable thanks to the exponential gains in l, which again can be derived from (1.20).

1.4.6. *Major arcs estimates*: "*low-low*" *case and* "*large scales*". At large scales, the major arcs become extremely narrow, so much so that the arithmetic frequencies at the center of these arcs can be given a common denominator Q with $\frac{1}{Q}$ much larger than the width of these arcs. In this regime it becomes possible to use a quantitative version of the Shannon sampling theorem (Theorem 4.18) to transfer from the integers $\mathbb{Z}$ to the adelic integers $\mathbb{A}_{\mathbb{Z}} = \mathbb{R} \times \hat{\mathbb{Z}}$ while essentially preserving all function space norms of interest. The behavior in the continuous variable $\mathbb{R}$ is relatively tractable due to the Ionescu–Wainger multiplier theory and [66]. The main difficulty is to understand the nature of the associated "arithmetic" average $A_{\hat{\mathbb{Z}}}$ on the profinite integers $\hat{\mathbb{Z}}$, which is a compact commutative ring. By some use of p-adic methods (see Appendix C), we will obtain a non-trivial L^p-improving estimate for this average, while from yet another invocation of Theorem 5.12 we will also obtain exponential decay in l for these averages (for the L^2 theory at least, and the remaining cases can then be treated by interpolation). By combining these estimates with some general manipulation of variational norms, and also relying primarily on a vector-valued version of Lépingle's inequality from [69] to handle the variational behavior in the continuous variable $\mathbb{R}$, we can obtain acceptable control on the contribution of the large scales.

1.4.7. *Major arcs estimates*: *remaining cases*. The other cases ("high-high," "low-high," "high-low") can be treated by modifications of the method; the main new difficulties are to obtain some additional decay when one is relatively far from the arithmetic frequencies at the center of the major arcs (that is to say, when the continuous component of the frequency is large). By interpolation one only needs to obtain this decay for the ℓ^2 theory. In the "high-high" case one can obtain such a decay using Theorem 5.12 once again, exploiting almost orthogonality in order to sum over scales N. In the remaining "low-high" and "high-low" cases we will obtain the required decay by applying an elementary integration by parts to a certain bilinear symbol associated to the averaging operation A_N (see Lemma 7.26). On the other hand, this decay is at risk of being overwhelmed by the increased oscillations present in the symbol. To avoid this we use shifted Calderón–Zygmund theory (see Appendix B), of the type used for instance in [63], that allows one to handle certain types of oscillating Fourier multipliers losing only acceptable logarithmic factors in the estimates. The idea of shifted maximal estimates

was also recently exploited in [46] in the context of establishing pointwise ergodic theorems for the polynomial averages on nilpotent groups; it seems to be decisive in problems when the operators in question cannot be interpreted as convolution operators corresponding to an abelian convolution.

1.4.8. *Final remarks.* Finally, we emphasize that the proof of Theorem 1.17 can also be adapted (and simplified) to give an alternate proof of Theorem 1.8 (but in which one only controls the long variation rather than the full variation). We sketch the changes needed to the argument as follows. The exponent p_1 is now fixed to equal ∞ (so that $p = p_2$), and the first function f is fixed to equal 1 (which allows for several simplifications, for instance the parameter l_1 can be taken to be 0, and s_1 can be taken to be $-u$). All appearances of $\mathbb{1}_{p_1=p_2=2}$ are now replaced by $\mathbb{1}_{p_1=\infty,p_2=2}$. Various linear estimates, such as Ionescu–Wainger multiplier estimates, shifted Calderón–Zygmund estimates, and Lepingle's inequality, do not hold in general at the ℓ^∞ endpoint, but are trivially true when applied to the specific function $f = 1$ in ℓ^∞, so this does not cause difficulty. Theorem 5.12 needs to be modified to an $\ell^\infty \times \ell^2 \to \ell^2$ estimate with $f = 1$, but in this case the required gain of $2^{-cl} + \langle \operatorname{Log} N\rangle^{-cC_1}$ is immediate from Plancherel's theorem and Weyl sum estimates [48, Lemma 20.3, p. 462], thus avoiding the need to invoke the Peluse–Prendiville theory.

1.5. *Open questions.* While our main interest is in averaging operators on the integers $\mathbb{Z}$, in the course of our arguments it became natural to also consider the analogous averaging operators on other locally compact abelian domains such as $\mathbb{R}, \mathbb{Z}/Q\mathbb{Z}, \mathbb{R} \times \mathbb{Z}/Q\mathbb{Z}, \mathbb{Z}/p^j\mathbb{Z}, \mathbb{Z}_p, \hat{\mathbb{Z}}$, and $\mathbb{A}_\mathbb{Z}$, with the adelic integers $\mathbb{A}_\mathbb{Z}$ playing a particularly central role, at least on a conceptual level; see Figure 1. The connection can be summarized by the slogan

$$\text{Major arc analysis on } \mathbb{Z} \approx \text{Low frequency analysis on } \mathbb{A}_\mathbb{Z},$$

where "low frequency" has to be interpreted in both a continuous and arithmetic sense; see Figure 7. In particular, the adelic averaging operators $A_{N,\mathbb{A}_\mathbb{Z}}$ defined in (7.12) emerge as a simplified model for the integer averaging operators $A_{N,\mathbb{Z}}$, and further investigations into similar problems in discrete harmonic analysis may wish to begin by first understanding adelic models of such problems, particularly in "true complexity zero" situations in which one suspects that the major arc contributions are dominant or equivalently that the minor arc contribution is negligible. In fact, the method of proof of Theorem 1.17 relies in an essential way on the negligibility of the minor arc contribution; in the language of additive combinatorics, this reflects the fact that the pattern $(x, x - n, x - P(n))$ has "true complexity zero" in the sense of Gowers and Wolf [39]. In the language of ergodic theory, the corresponding assertion is that the minimal characteristic factor of the averages $A_N^{\mathrm{n},P(\mathrm{n})}$ is the rational Kronecker (or profinite) factor $\mathcal{K}_{\mathrm{rat}}$ generated by the periodic functions.

We close our introduction with some questions relating to Theorem 1.17 that remain open.

(1) Does Theorem 1.17 continue to hold if one of p_1, p_2 is allowed to be infinite? Certainly from Theorem 1.8 the maximal inequality (Theorem 1.17(iii)) will still hold if one or both of p_1, p_2 are infinite, but the situation for the other parts of the theorem are less clear (except in the special case where $p_1 = \infty$ and f is constant, or $p_2 = \infty$ and g is constant). Given the ability to break duality, the endpoints $p_1 = 1$, $p_2 = 1$ could also be investigated.

(2) Is the analogue of Theorem 1.17(iv) true for the full variation, in which the lacunarity hypothesis on $\mathbb{D}$ is omitted? Equivalently, can the implied constant in (1.18) be made uniform in λ? The problem is likely to be significantly simpler if the sharp truncation $\mathbb{1}_{n \leq N}$ implicit in the definition of the averages $A_N^{\mathrm{n}, P(\mathrm{n})}$ is replaced by a smoother weight. Note that the linear analogue of this question was already resolved in Theorem 1.8(iv).

(3) To what extent can the results in Theorem 1.17 extend to other bilinear averages $A_N^{P_1(\mathrm{n}), P_2(\mathrm{n})}$, or more ambitiously to general multilinear polynomial ergodic averages $A_N^{P_1(\mathrm{n}), \ldots, P_k(\mathrm{n})}$? We refer to Bergelson's surveys [3, Question 9, pp. 52], [4, §6, pp. 838]. It is not difficult to adapt Theorem 1.17 to cover averages $A_N^{P_1(\mathrm{n}), P_2(\mathrm{n})}$ in which one of the P_1, P_2 is linear (i.e., of degree 1) and the other is non-linear, however when both P_1, P_2 are non-linear a refinement of the Peluse–Prendiville theory may be required. We hope to investigate these averages in future work.

(4) Is there some analogue of these methods that can cover patterns of higher complexity? A natural first step would be to recover some portion of Theorem 1.16 (which has "true complexity one" in the Gowers–Wolf [39] sense) by these methods.

(5) What are explicit ranges of exponents p_1, p_2 for which one can "break duality" with in Theorem 1.17? In the model case $p_1 = p_2$ (so that $p = p_1/2 = p_2/2$), Lemma 11.1 suggests that one should be able to take p in the range $p > 1 - \frac{1}{d^2+d-1}$, or even $p > 1 - \frac{1}{2d}$ in the $d = 2$ case, with the latter range also expected if [41, Conj. 1.5] holds. It should also be possible to recover the optimal range $r > 2$ of the variational exponent r below the line of duality. (Our current arguments incur a loss in this parameter that depends on how close $(1/p_1, 1/p_2)$ is to $(1/2, 1/2)$.)

(6) Theorem 1.17(iv) gives variational estimates in $\boldsymbol{V}^r$ norms for $r > 2$, and in Section 12 the $r = 2$ endpoint is shown to be false. However, there still remains the question of whether a jump inequality (analogous to Doob's inequality for martingales) is true at the $r = 2$ endpoint. Such endpoint jump inequalities were established in [68] for linear polynomial averages on $\mathbb{Z}^k$.

(7) Theorem 1.17 was focused on unweighted averages

$$A_N^{\mathrm{n},P(\mathrm{n})}(f,g)(x) = \mathbb{E}_{n\in[N]} T^n f(x) T^{P(n)} g(x),$$

but one can pose similar questions[3] for the truncated singular integral analogue $\sum_{0<|n|<N} \frac{1}{n} T^n f(x) T^{P(n)} g(x)$; currently only single-scale super-Hölder estimates are known [26]. In the linear setting (resp. the bilinear setting for two linear polynomials), the theory for the averages and the truncated singular integrals are similar; see [68] (resp. [55]). Bounds on the (untruncated) bilinear continuous singular integrals were obtained in [60], [62], [61], [63].

(8) To what extent do the implied constants in Theorem 1.17 depend on the coefficients of P? The estimate in [41, Th. 1.6] suggests that the dependence of constants is at worst polynomial; on the other hand, [68, Cor. 1.15] suggests that one may be able to obtain bounds uniform in the coefficients, by lifting the problem to $\mathbb{Z}^d$ and establishing an analogue of Theorem 1.17 in that setting. (However, this latter strategy would require a multidimensional version of the theory of Peluse and Prendiville, which may be highly nontrivial.) We also hope to investigate the latter multidimensional strategy in future work.

(9) Can the results here on the (rational) integers $\mathbb{Z}$ be extended to rings of integers in more general number fields, such as the ring $\mathbb{Z}[i]$ of Gaussian integers? Certainly the adelic formalism is exceptionally well adapted to this setting [81], but other components of the argument may require significantly more effort to generalize appropriately.

(10) Assuming that $P \in \mathbb{R}[\mathrm{n}]$, it also makes sense to ask whether Theorem 1.17 holds with the averages $A_N^{\mathrm{n},\lfloor P(\mathrm{n})\rfloor}(f,g)$ in place of $A_N^{\mathrm{n},P(\mathrm{n})}(f,g)$. This kind of question for linear polynomial averages was considered by Bourgain in [13]. One could also replace the polynomial P with elements of other Hardy fields, in the spirit of [8], [9], or by random functions of polynomial growth, in the spirit of [32]. In fact these variants may be simpler than the polynomial case, as the only major arc that is expected to be significant is the one centered at the origin.

(11) As mentioned previously, there is a well-developed theory of characteristic factors for the limiting values of non-conventional polynomial averages $A_N(f_1,\dots,f_k)$ when the functions $f_1,\dots,f_k$ lie in $L^\infty(X)$ and X has finite

[3] One could also consider fractional integral type expressions

$$\mathbb{E}_{n\in[N]} (n/N)^{-\alpha} T^n f(x) T^{P(n)} g(x)$$

for $0 < \alpha < 1$, but these can be easily expressed as linear combinations of the unweighted averages via summation by parts and so would be expected to obey nearly identical estimates to those averages.

measure; see [3], [4], [30]. To what extent does this theory extend to other L^p spaces and to the case when X is merely σ-finite, for instance for the average $A_N^{\mathrm{n},P(\mathrm{n})}(f,g)$ studied in Theorem 1.17?

Acknowledgments. We thank Sarah Peluse and Sean Prendiville for several helpful discussions about their inverse theory and for sharing some unpublished notes. We also thank Jim Wright for sharing his unpublished notes on L^q-improving estimates in p-adic fields $\mathbb{Q}_p$, and for helpful comments and corrections. We thank Vitaly Bergelson and Bryna Kra for the discussion about the history of Problem 11 from Frantzikinakis' open problems survey [30]. We also thank Jaume de Dios and Dariusz Kosz and Wojciech Słomian for further corrections. Finally, we thank the referees for careful reading of the manuscript and useful remarks that led to the improvement of the presentation.

2. Notation

In this section we set out some basic notation used throughout the paper.

2.1. *Elementary number theory.* We use $\mathbb{Z}_+ := \{1,2,\dots\}$ to denote the positive integers and $\mathbb{N} := \{0,1,2,\dots\}$ to denote the natural numbers. For any $N > 0$, $[N]$ denotes the discrete interval $[N] := \{n \in \mathbb{Z}_+ : n \le N\}$. The set $\{2,3,5,\dots\}$ of all prime numbers will be denoted by $\mathbb{P}$. If $q_1, q_2 \in \mathbb{Z}_+$, we write $q_1|q_2$ if q_1 divides q_2. If $a, q \in \mathbb{Z}_+$, we let (a,q) denote the greatest common divisor of a and q. We let $[q]^\times := \{a \in [q] : (a,q) = 1\}$ denote the elements of $[q]$ that are coprime to q.

2.2. *Magnitudes and asymptotic notation.* We use the Japanese bracket notation

$$\langle x \rangle := (1 + |x|^2)^{1/2}$$

for any real or complex x. We use $\lfloor x \rfloor$ to denote the greatest integer less than or equal to x. All logarithms in this paper will be to base 2, and for any $N \ge 1$, we define the *logarithmic scale* $\operatorname{Log} N$ of N by the formula

$$\operatorname{Log} N := \lfloor \log N \rfloor, \tag{2.1}$$

thus $\operatorname{Log} N$ is the unique natural number such that $2^{\operatorname{Log} N} \le N < 2^{\operatorname{Log} N + 1}$.

For any two quantities A, B, we will write $A \lesssim B$, $B \gtrsim A$, or $A = O(B)$ to denote the bound $|A| \le CB$ for some absolute constant C. If we need the implied constant C to depend on additional parameters, we will denote this by subscripts; thus, for instance, $A \lesssim_\rho B$ denotes the bound $|A| \le C_\rho B$ for some C_ρ depending on ρ. We write $A \sim B$ for $A \lesssim B \lesssim A$. To abbreviate the notation we will sometimes explicitly permit the implied constant to depend on certain fixed parameters (such as the polynomial P) when the issue of uniformity with respect to such parameters is not of relevance.

2.3. *Averages, indicators, and cutoffs.* We use the averaging notation

$$\mathbb{E}_{n\in A} f(n) := \frac{1}{\#A}\sum_{n\in A} f(n) \tag{2.2}$$

for any finite non-empty set A, where $\#A$ denotes the cardinality of A; in other words, $\mathbb{E}_{n\in A} f(n)$ is the integral of f against normalized counting measure on A. Note, in particular, that $\mathbb{E}_{n\in[N]} f(n) = \frac{1}{N}\sum_{n=1}^{N} f(n)$ when $N \in \mathbb{Z}_+$. We use $\mathbb{1}_E$ to denote the indicator function of a set E. Similarly, if S is a statement, we use $\mathbb{1}_S$ to denote its indicator, equal to 1 if S is true and 0 if S is false. Thus, for instance, $\mathbb{1}_E(x) = \mathbb{1}_{x\in E}$.

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 $[-1/2,1/2]$. All constants are permitted to depend on η. For any $k \in \mathbb{Z}$, we let $\eta_{\leq k}\colon \mathbb{R} \to [0,1]$ denote the rescaled version of η:

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

2.4. *Function spaces.* All vector spaces in this paper will be over the complex numbers $\mathbb{C}$.

If $T\colon V \to W$ is a continuous linear map between normed vector spaces V, W, we use $\|T\|_{V\to W}$ to denote its operator norm. If $B\colon V_1 \times V_2 \to W$ is a continuous bilinear map between normed vector spaces V_1, V_2, W, we similarly use $\|B\|_{V_1\times V_2\to W}$ to denote its operator norm.

If (X,μ) is a measure space, we let $L^0(X)$ be the space of all μ-measurable complex-valued functions defined on X, with the usual convention of identifying functions that agree μ-almost everywhere. The space of all functions in $L^0(X)$ whose modulus is integrable with p-th power is denoted by $L^p(X)$ for $p \in (0,\infty)$, whereas $L^\infty(X)$ denotes the space of all essentially bounded functions in $L^0(X)$. If $1 \leq p \leq \infty$ is an exponent, the dual exponent $1 \leq p' \leq \infty$ is defined by the usual relation $1/p + 1/p' = 1$. When X is endowed with counting measure, we will abbreviate $L^p(X)$ to $\ell^p(X)$ or even ℓ^p.

We can extend these notions to functions taking values in a finite dimensional normed vector space $V = (V, \|\cdot\|_V)$, for instance $L^0(X;V)$ is the space of measurable functions from X to V (up to almost everywhere equivalence), and

$$L^p(X;V) := \left\{F \in L^0(X;V) : \|F\|_{L^p(X;V)} := \|\|F\|_V\|_{L^p(X)} < \infty\right\}. \tag{2.3}$$

One can extend these notions to infinite-dimensional V, at least if V is separable, but we will almost always be able to work in finite-dimensional settings (or we can quickly reduce to such a setting by a standard approximation argument).

For any finite dimensional normed vector space $(B, \|\cdot\|_B)$ and any sequence $(\mathfrak{a}_t)_{t\in\mathbb{I}}$ of elements of B indexed by a totally ordered set $\mathbb{I}$, and any exponent

$1 \le r < \infty$, the r-variation seminorm is defined by the formula

$$\|(\mathfrak{a}_t)_{t\in\mathbb{I}}\|_{V^r(\mathbb{I};B)} := \sup_{J\in\mathbb{Z}_+} \sup_{\substack{t_0\le\cdots\le t_J\\ t_j\in\mathbb{I}}} \Big(\sum_{j=0}^{J-1} \|\mathfrak{a}(t_{j+1}) - \mathfrak{a}(t_j)\|_B^r\Big)^{1/r}, \tag{2.4}$$

where the supremum is taken over all finite increasing sequences in $\mathbb{I}$, and it is set by convention to equal zero if $\mathbb{I}$ is empty. Taking limits as $r \to \infty$ we also adopt the convention

$$\|(\mathfrak{a}_t)_{t\in\mathbb{I}}\|_{V^\infty(\mathbb{I};B)} := \sup_{t\le t'\in\mathbb{I}} \|\mathfrak{a}(t') - \mathfrak{a}(t)\|_B.$$

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

$$\|(\mathfrak{a}_t)_{t\in\mathbb{I}}\|_{\boldsymbol{V}^r(\mathbb{I};B)} := \sup_{t\in\mathbb{I}} \|\mathfrak{a}_t\|_B + \|(\mathfrak{a}_t)_{t\in\mathbb{I}}\|_{V^r(\mathbb{I};B)}. \tag{2.5}$$

This clearly defines a norm on the space of functions from $\mathbb{I}$ to B. If $B = \mathbb{C}$, then we will abbreviate $V^r(\mathbb{I};X)$ to $V^r(\mathbb{I})$ or V^r, and $\boldsymbol{V}^r(\mathbb{I};X)$ to $\boldsymbol{V}^r(\mathbb{I})$ or $\boldsymbol{V}^r$. If (X,μ) is a measure space, then using (2.5) and (2.3), one can explicitly write

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

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

$$\|(\mathfrak{a}_t)_{t\in\mathbb{I}}\|_{\boldsymbol{V}^r(\mathbb{I};X)} \lesssim \|(\mathfrak{a}_t)_{t\in\mathbb{I}_1}\|_{\boldsymbol{V}^r(\mathbb{I}_1;X)} + \|(\mathfrak{a}_t)_{t\in\mathbb{I}_2}\|_{\boldsymbol{V}^r(\mathbb{I}_2;X)} \tag{2.6}$$

whenever $\mathbb{I} = \mathbb{I}_1 \uplus \mathbb{I}_2$ is an ordered partition of $\mathbb{I}$, thus $t_1 < t_2$ for all $t_1 \in \mathbb{I}_1, t_2 \in \mathbb{I}_2$. In a similar spirit we have the bound

$$\|(\mathfrak{a}_t)_{t\in\mathbb{I}}\|_{\boldsymbol{V}^r(\mathbb{I};X)} \lesssim \|(\mathfrak{a}_t)_{t\in\mathbb{I}}\|_{\ell^r(\mathbb{I};X)} \le \|(\mathfrak{a}_t)_{t\in\mathbb{I}}\|_{\ell^1(\mathbb{I};X)}. \tag{2.7}$$

From Hölder's inequality one easily establishes the algebra property

$$\|(\mathfrak{a}_t\mathfrak{b}_t)_{t\in\mathbb{I}}\|_{\boldsymbol{V}^r} \lesssim \|(\mathfrak{a}_t)_{t\in\mathbb{I}}\|_{\boldsymbol{V}^r} \|(\mathfrak{b}_t)_{t\in\mathbb{I}}\|_{\boldsymbol{V}^r} \tag{2.8}$$

for any scalar sequences $(\mathfrak{a}_t)_{t\in\mathbb{I}}$, $(\mathfrak{b}_t)_{t\in\mathbb{I}}$.

2.5. *Tensor products.* Given two functions $f\colon X \to \mathbb{C}$, $g\colon Y \to \mathbb{C}$, we define their tensor product $f \otimes g\colon X \to Y \to \mathbb{C}$ by the formula

$$f \otimes g(x,y) := f(x)g(y).$$

One can also define the formal tensor product $f \otimes g$ of elements $f \in V$, $g \in W$ of abstract vector spaces V, W, which takes values in the algebraic tensor product $V \otimes W$. By abuse of notation, we identify these two notions of tensor product.

If $T_1\colon V_1 \to W_1$, $T_2\colon V_2 \to W_2$ are linear maps, we define the tensor product $T_1 \otimes T_2\colon V_1 \otimes V_2 \to W_1 \otimes W_2$ as the unique linear map such that

$$T_1 \otimes T_2(f_1 \otimes f_2) = (T_1 f_1) \otimes (T_2 f_2) \tag{2.9}$$

whenever $f_1 \in V_1$, $f_2 \in V_2$. Similarly, if $B_1\colon U_1 \times V_1 \to W_1$ and $B_2\colon U_2 \times V_2 \to W_2$ are bilinear maps, we define $B_1 \otimes B_2\colon (U_1 \otimes U_2) \times (V_1 \otimes V_2) \to W_1 \otimes W_2$ to be the unique bilinear map such that

$$B_1 \otimes B_2(f_1 \otimes f_2, g_1 \otimes g_2) = B_1(f_1, g_1) \otimes B_2(f_2, g_2) \tag{2.10}$$

whenever $f_1 \in U_1, g_1 \in V_1, f_2 \in U_2, g_2 \in V_2$. This algebraic tensor product can often be extended to analytic settings. For instance, if $T_1 \colon L^p(X_1) \to L^q(Y_1)$ and $T_2 \colon L^p(X_2) \to L^q(Y_2)$ are integral operators of the form

$$T_1 f_1(y_1) = \int_{X_1} K_1(x_1, y_1) f_1(x_1)\, d\mu_{X_1}(x_1)$$

and

$$T_2 f_2(y_2) = \int_{X_2} K_2(x_2, y_2) f_2(x_2)\, d\mu_{X_2}(x_2)$$

then one can define $T_1 \otimes T_2 \colon L^p(X_1 \times X_2) \to L^q(Y_1 \times Y_2)$ (formally, at least) by

$$(T_1 \otimes T_2) f(y_1, y_2) = \int_{X_1 \times X_2} K_1(x_1, y_1) K_2(x_2, y_2) f(x_1, x_2)\, d\mu_{X_1}(x_1) d\mu_{X_2}(x_2).$$

We claim the multiplicativity property

$$\|T_1 \otimes T_2\|_{L^p(X_1 \times X_2) \to L^q(Y_1 \times Y_2)} = \|T_1\|_{L^p(X_1) \to L^q(Y_1)} \|T_2\|_{L^p(X_2) \to L^q(Y_2)}, \tag{2.11}$$

in the case[4] where one of the kernels (say K_1) is non-negative, and assuming X_1, X_2, Y_1, Y_2 are σ-finite with positive measure to avoid degeneracies, by the following argument. The lower bound is clear by testing $T_1 \otimes T_2$ on tensor products $f_1 \otimes f_2$, so we focus on the upper bound (which is what is needed in our applications). If $f \in L^p(X_1 \times X_2)$, we have

$$(T_1 \otimes T_2) f(y_1, y_2) = \int_{X_1} K_1(x_1, y_1) T_2(f_{x_1})(y_2)\, d\mu_{X_1}(x_1),$$

where $f_{x_1} \colon x_2 \mapsto f(x_1, x_2)$ denotes the slice of f, hence for any $y_1 \in Y_1$ and by the non-negativity of K_1, we have

$$\|(T_1 \otimes T_2) f(y_1, \cdot)\|_{L^q(Y_2)} \leq \|T_2\|_{L^p(X_2) \to L^q(Y_2)} \int_{X_1} K_1(x_1, y_1) \|f_{y_1}\|_{L^p(X_1)}\, d\mu_{X_1}(x_1).$$

Taking $L^q(Y_1)$ norms of both sides and using the Fubini–Tonelli theorem, we conclude that

$$\|(T_1 \otimes T_2) f\|_{L^q(Y_1 \times Y_2)} \leq \|T_1\|_{L^p(X_1) \to L^q(Y_1)} \|T_2\|_{L^p(X_2) \to L^q(Y_2)} \|f\|_{L^p(X_1 \times X_2)},$$

[4]There is another case where (2.11) holds, namely when $q \geq p$ and no non-negativity hypothesis is assumed, by factoring $T_1 \otimes T_2 = (T_1 \otimes \mathrm{id}) \circ \mathrm{id} \circ (\mathrm{id} \otimes T_2)$ and establishing the inequalities

$$\|\mathrm{id} \otimes T_2\|_{L^p(X_1 \times X_2) \to L^p(X_1; L^q(Y_2))} \leq \|T_2\|_{L^p(X_2) \to L^q(Y_2)},$$
$$\|\mathrm{id}\|_{L^p(X_1; L^q(Y_2)) \to L^q(Y_2; L^p(X_1))} \leq 1,$$

and

$$\|T_1 \otimes \mathrm{id}\|_{L^q(Y_2; L^p(X_1)) \to L^q(Y_1 \times Y_2)} \leq \|T_1\|_{L^p(X_1) \to L^q(Y_1)}.$$

However, this argument does not easily extend to the bilinear case, which is the case of most interest to us.

giving the claim. An analogous argument gives the identity

$$\|B_1 \otimes B_2\|_{L^p(X_1\times X_2)\times L^q(Y_1\times Y_2)\to L^r(Z_1\times Z_2)} = \|B_1\|_{L^p(X_1)\times L^q(Y_1)\to L^r(Z_1)}\|B_2\|_{L^p(X_2)\times L^q(Y_2)\to L^r(Z_2)} \tag{2.12}$$

for tensor products of bilinear operators, with (say) B_1 arising from a non-negative kernel, again assuming all spaces σ-finite with positive measure to avoid degeneracies.

3. **Transferring to the integer shift**

In this section we perform three standard and general reductions for our problem:

(i) By standard limiting arguments, we show that long variational estimates, such as the one in Theorem 1.17(iv), are sufficient to establish maximal inequalities, norm convergence, and pointwise almost everywhere convergence. Thus we can focus exclusively on variational estimates in the sequel.

(ii) We apply the Calderón transference principle (see, e.g., [23, App. A]) to transfer the long variational estimates to the integer shift system $\mathbb{Z} = (\mathbb{Z}, \mu_{\mathbb{Z}}, T_{\mathbb{Z}})$. As mentioned in the introduction, this allows us to exploit the Fourier-analytic structure of $\mathbb{Z}$ (and eventually, $\mathbb{A}_{\mathbb{Z}}$ as well).

(iii) We use a telescoping argument to replace the averaging operator

$$A_N^{P_1,\dots,P_k}(f_1,\dots,f_k)(x) = \mathbb{E}_{n\in[N]} f_1(T^{P_1(n)}x)\cdots f_k(T^{P_k(n)}x)$$

with the upper half[5]

$$\tilde{A}_N^{P_1,\dots,P_k}(f_1,\dots,f_k)(x) = \mathbb{E}_{n\in[N]} f_1(T^{P_1(n)}x)\cdots f_k(T^{P_k(n)}x)\mathbb{1}_{n>N/2}. \tag{3.1}$$

This technical reduction is convenient as it allows one to avoid the stationary points of the polynomials $P_1,\dots,P_k$. (In particular, we get good lower bounds on the first derivatives of these polynomials.)

These reductions are available for arbitrary non-conventional averages, not just for the bilinear averages $A_N^{\mathrm{n},P(\mathrm{n})}$ treated in this paper, so we give these reductions in the general setting.

Proposition 3.2 (General reductions). *Let (X,μ,T) be a measure-preserving system, let $P_1(\mathrm{n}),\dots,P_k(\mathrm{n}) \in \mathbb{Z}[\mathrm{n}]$, let $0 < p_1,\dots,p_k,p < \infty$, and let $1 \le r < \infty$.*

[5]One could also work with the normalized upper half $\frac{\lfloor N\rfloor}{\lfloor N/2\rfloor}\tilde{A}_N$ here if desired, though it makes little difference to the subsequent arguments other than adjusting a few constants by a factor of two.

(i) (Reduction to variational estimate). *Suppose one has the variational estimate*

$$\begin{aligned}\|(A_N^{P_1,\dots,P_k}(f_1,\dots,f_k))_{N\in\mathbb{D}}\|_{L^p(X;\boldsymbol{V}^r)}\\ \lesssim_{p_1,\dots,p_k,p,P_1,\dots,P_k,r,\lambda} \|f_1\|_{L^{p_1}(X)}\cdots\|f_k\|_{L^{p_k}(X)}\end{aligned} \tag{3.3}$$

for all $\lambda > 1$ and $f_i \in L^{p_i}(X)$, $i = 1,\dots,k$, and all finite λ-lacunary subsets $\mathbb{D}$ of $[1,+\infty)$. Then one has the maximal inequality

$$\begin{aligned}\|(A_N^{P_1,\dots,P_k}(f_1,\dots,f_k))_{N\in\mathbb{Z}_+}\|_{L^p(X;\ell^\infty)}\\ \lesssim_{p_1,\dots,p_k,p,P_1,\dots,P_k,r} \|f_1\|_{L^{p_1}(X)}\cdots\|f_k\|_{L^{p_k}(X)},\end{aligned} \tag{3.4}$$

and for any $f_i \in L^{p_i}(X)$, $i = 1,\dots,k$, the averages $A_N^{P_1,\dots,P_k}(f_1,\dots,f_k)$ converge pointwise almost everywhere and in $L^p(X)$ norm.

(ii) (Calderón transference principle). *Suppose that we are in the Hölder exponent case $\frac{1}{p_1}+\cdots+\frac{1}{p_k}=\frac{1}{p}$. Then in order to establish (3.3) for arbitrary measure-preserving systems $X = (X,\mu,T)$, it suffices to show (3.3) for the integer shift model $\mathbb{Z} = (\mathbb{Z},\mu_{\mathbb{Z}},T_{\mathbb{Z}})$.*

(iii) (Telescoping argument). *In order to establish (3.3) under the assumptions of* (i), *it suffices to establish the bound*

$$\begin{aligned}\|(\tilde{A}_N^{P_1,\dots,P_k}(f_1,\dots,f_k))_{N\in\mathbb{D}}\|_{L^p(X;\boldsymbol{V}^r)}\\ \lesssim_{p_1,\dots,p_k,p,P_1,\dots,P_k,r,\lambda} \|f_1\|_{L^{p_1}(X)}\cdots\|f_k\|_{L^{p_k}(X)}\end{aligned} \tag{3.5}$$

under the same assumptions, where $\tilde{A}_n^{P_1,\dots,P_k}$ is defined in (3.1).

Note that all of the reductions in this proposition apply in both the Banach exponent case $p \geq 1$ and the non-Banach exponent case $0 < p < 1$. However, we emphasize that the Calderón transference principle (ii) is only available in the Hölder exponent case $\frac{1}{p_1}+\cdots+\frac{1}{p_k}=\frac{1}{p}$.

Proof. To simplify the notation we allow all implied constants to depend on $p_1,\dots,p_k,P_1,\dots,P_k,r$.

We begin with (i). Fix $f_1,\dots,f_k$, and abbreviate $A_N^{P_1,\dots,P_k}(f_1,\dots,f_k)(x)$ as $a_N(x)$ for any $N \geq 1$. For any $s \in \mathbb{Z}_+$, introduce the $2^{1/s}$-lacunary set

$$2^{\mathbb{N}/s} := \{2^{n/s} : n \in \mathbb{N}\}.$$

(Note that here we exploit the freedom to choose scales N that are real-valued rather than integer-valued). From (3.3) and monotone convergence, we have

$$\|(a_N)_{N\in 2^{\mathbb{N}/s}}\|_{L^p(X;\boldsymbol{V}^r)} \lesssim_s \|f_1\|_{L^{p_1}(X)}\cdots\|f_k\|_{L^{p_k}(X)}. \tag{3.6}$$

To prove (3.4), we may assume without loss of generality that $f_1,\dots,f_k$ are non-negative, thanks to the pointwise triangle inequality

$$|A_N^{P_1,\dots,P_k}(f_1,\dots,f_k)| \leq A_N^{P_1,\dots,P_k}(|f_1|,\dots,|f_k|).$$

In the non-negative case we have the additional pointwise bound

$$\sup_{N\in\mathbb{Z}_+} a_N(x) \le 2 \sup_{N\in 2^{\mathbb{N}}} a_N(x),$$

and the claim (3.4) now follows from (3.6).

Now we establish pointwise convergence. By linearity we may assume that the $f_1,\dots,f_k$ are all non-negative. From (3.6) and (3.4), we see that for almost all $x\in X$, the quantity

$$M(x) := \sup_{N\in\mathbb{Z}_+} a_N(x)$$

is finite, as are the variational norms $\|(a_N)_{N\in 2^{\mathbb{N}/s}}\|_{\boldsymbol{V}^r}$ for every $s\in\mathbb{Z}_+$. From the latter we conclude that the limits $\lim_{N\to\infty;N\in 2^{\mathbb{N}/s}} a_N(x)$ exist almost everywhere for all $s\ge 1$; since $2^{\mathbb{N}}\subset 2^{\mathbb{N}/s}$, this limit is independent of s, thus

$$\lim_{N\to\infty;N\in 2^{\mathbb{N}/s}} a_N(x) = a_\infty(x)$$

for some $a_\infty(x)$. For any sufficiently large N, if we let N' be the first element of $2^{\mathbb{N}/s}$ greater than or equal to N, we see from the triangle inequality that

$$a_N(x) = a_{N'}(x) + O(M(x)/s).$$

Hence on taking limits

$$\liminf_{N\to\infty} a_N(x), \limsup_{N\to\infty} a_N(x) = a_\infty(x) + O(M(x)/s),$$

sending $s\to\infty$, we conclude that $a_N(x)$ converges to $a_\infty(x)$ as $N\to\infty$ as claimed. Finally, norm convergence follows from pointwise convergence, the maximal inequality, and the dominated convergence theorem. This proves (i).

Now we prove (ii). This follows from the general Calderón transference principle [18], but for the convenience of the reader we supply a proof here. We first observe from the Fubini–Tonelli theorem and Hölder's inequality (and the Hölder exponent hypothesis $\frac{1}{p_1}+\cdots+\frac{1}{p_k}=\frac{1}{p}$) that if (3.3) is established for the integer shift model $(\mathbb{Z},\mu_{\mathbb{Z}},T_{\mathbb{Z}})$, then it automatically holds for any product system $(X\times\mathbb{Z},\mu\times\mu_{\mathbb{Z}},\mathrm{id}\times T_{\mathbb{Z}})$, where (X,μ) is an arbitrary σ-finite measure space and $\mathrm{id}\times T_{\mathbb{Z}}$ is the shift $(x,n)\mapsto(x,n-1)$, since there is no interaction between the individual fibers $\{x\}\times\mathbb{Z}, x\in X$ of this system.

Now let (X,μ,T) be an arbitrary measure-preserving system. To prove (3.3), it suffices by multilinearity to do so when the f_i are non-negative. We may assume that each of the f_i are bounded and supported on a set of finite measure. We may normalize $\|f_i\|_{L^{p_i}(X)}=1$ for $i=1,\dots,k$, thus our task is now to show that

$$\|(A_{N,X}^{P_1,\dots,P_k}(f_1,\dots,f_k))_{N\in\mathbb{D}}\|_{L^p(X;\boldsymbol{V}^r)} \lesssim_\lambda 1.$$

Now let M be a large natural number, let $D := \max_{i\in[k]}\deg P_i$, and let $C>0$ be a quantity to be specified later that can depend on $\mathbb{D},P_1,\dots,P_k$ but is

independent of M. On the product system $X \times \mathbb{Z} = (X \times \mathbb{Z}, \mu \times \mu_{\mathbb{Z}}, \mathrm{id} \times T_{\mathbb{Z}})$, define the functions

$$f_{i,M}(x,n) := \mathbb{1}_{[3CM^D]}(n) f_i(T^{-n}x)$$

for $i = 1, \dots, k$. From the Fubini–Tonelli theorem and the measure-preserving nature of T, one has

$$\|f_{i,M}\|_{L^{p_i}(X\times\mathbb{Z})} = (3CM)^{D/p_i}.$$

Also, we observe the identity

$$\begin{aligned}\|(A^{P_1,\dots,P_k}_{N,X\times\mathbb{Z}}(f_{1,M},\dots,f_{k,M})(x,n))_{N\in\mathbb{D}\cap[M]}\|_{\boldsymbol{V}^r}\\ = \|(A^{P_1,\dots,P_k}_{N,X}(f_1,\dots,f_k)(T^{-n}x))_{N\in\mathbb{D}\cap[M]}\|_{\boldsymbol{V}^r}\end{aligned}$$

whenever $CM^D \le n \le 2CM^D$. From the Fubini–Tonelli theorem again, we conclude that

$$\begin{aligned}\|(A^{P_1,\dots,P_k}_{N,X\times\mathbb{Z}}(f_{1,M},\dots,f_{k,M}))_{N\in\mathbb{D}}\|_{L^p(X\times\mathbb{Z};\boldsymbol{V}^r)}\\ \ge (CM^D+1)^{1/p}\|(A^{P_1,\dots,P_k}_{N,X}(f_1,\dots,f_k))_{N\in\mathbb{D}\cap[M]}\|_{L^p(X;\boldsymbol{V}^r)}.\end{aligned}$$

Applying (3.3) to the product system $X \times \mathbb{Z}$, we conclude that

$$\|(A^{P_1,\dots,P_k}_{N,X}(f_1,\dots,f_k))_{N\in\mathbb{D}\cap[M]}\|_{L^p(X;\boldsymbol{V}^r)} \lesssim_\lambda M^{-D/p} M^{D/p_1} \dots M^{D/p_k};$$

using the Hölder exponent hypothesis $1/p_1 + \dots + 1/p_k = 1/p$ and sending $M \to \infty$, we obtain the claim.

Finally, we prove (iii). By linearity we may take $f_1, \dots, f_k$ to be nonnegative. Fix $\lambda > 1$, and set

$$\tilde{a}_N(x) := \tilde{A}^{P_1,\dots,P_k}_N(f_1,\dots,f_k)(x).$$

We observe the telescoping identity

$$a_N = \sum_{k=0}^{\infty} \frac{\lfloor N/2^k \rfloor}{\lfloor N \rfloor} \tilde{a}_{N/2^k} \mathbb{1}_{2^k \le N}.$$

We have $\frac{\lfloor N/2^k \rfloor}{\lfloor N \rfloor} = 2^{-k} + O(1/N)$, and hence by the triangle inequality we have the pointwise estimate

$$\|(a_N)_{N\in\mathbb{D}}\|_{\boldsymbol{V}^r} \le \sum_{k=0}^{\infty} 2^{-k} \|(\tilde{a}_{N/2^k}\mathbb{1}_{2^k\le N})_{N\in\mathbb{D}}\|_{\boldsymbol{V}^r} + O\Big(\sum_{k=0}^{\infty}\sum_{N\in\mathbb{D}} \frac{1}{N}\mathbb{1}_{2^k\le N}|\tilde{a}_{N/2^k}|\Big)$$

for all $x \in X$. Since the rescaling $\{N/2^k : N \in \mathbb{D}, 2^k \le N\}$ of a λ-lacunary set $\mathbb{D}$ is still λ-lacunary, we have from (3.5) that

$$\|(\tilde{a}_{N/2^k}\mathbb{1}_{2^k\le N})_{N\in\mathbb{D}}\|_{L^p(X;\boldsymbol{V}^r)} \lesssim_\lambda \|f_1\|_{L^{p_1}(X)} \cdots \|f_r\|_{L^{p_r}(X)}.$$

From (3.5) applied to singleton λ-lacunary sets, we have

$$\|\tilde{a}_{N/2^k}\|_{L^p(X)} \lesssim_\lambda \|f_1\|_{L^{p_1}(X)} \cdots \|f_r\|_{L^{p_r}(X)}.$$

Summing in N, k, using the triangle inequality $\|\sum_i f_i\|_{L^p(X)} \leq \sum_i \|f_i\|_{L^p(X)}$ (when $p \geq 1$) or the quasi-triangle inequality

$$\Big\| \sum_i f_i \Big\|_{L^p(X)}^p \leq \sum_i \|f_i\|_{L^p(X)}^p \tag{3.7}$$

(when $0 < p < 1$), we obtain the claim. □

Remark 3.8. A modification of the Calderón transference principle also allows us to handle measure-preserving systems in which the shift map T is not assumed to be invertible, as long as we also require the polynomials $P_1, \dots, P_k$ to be non-negative on $\mathbb{Z}_+$ so that the averaging operators $A_N^{P_1,\dots,P_k}$ remain well defined. We leave the details to the interested reader.

In view of this general proposition, Theorem 1.17 will now follow from

THEOREM 3.9 (Variational ergodic theorem on the integers). *Let $P(\mathrm{n}) \in \mathbb{Z}[\mathrm{n}]$ have degree $d \geq 2$, let $1 \leq p_1, p_2, p < \infty$ be such that $\frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{p}$, and let $f \in \ell^{p_1}(\mathbb{Z}), g \in \ell^{p_2}(\mathbb{Z})$. If $r > 2$ and $\lambda > 1$, then*

$$\|(\tilde{A}_N^{\mathrm{n},P(\mathrm{n})}(f,g))_{N\in\mathbb{D}}\|_{\ell^p(\mathbb{Z};\boldsymbol{V}^r)} \lesssim_{p_1,p_2,r,P,\lambda} \|f\|_{\ell^{p_1}(\mathbb{Z})}\|g\|_{\ell^{p_2}(\mathbb{Z})} \tag{3.10}$$

for all finite λ-lacunary subsets $\mathbb{D}$ of $[1, +\infty)$.

It remains to establish Theorem 3.9. This is the objective of much of the remainder of the paper.

Remark 3.11. It is essential in Theorem 3.9 for ergodic theory applications that one has the Hölder condition $\frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{p}$. In the super-Hölder regime $\frac{1}{p_1} + \frac{1}{p_2} > \frac{1}{p}$ it is easy to establish (3.10); for instance, when $(p_1, p_2, p) = (2, 2, \infty)$, it follows from Cauchy-Schwarz that

$$\|\tilde{A}_N^{\mathrm{n},P(\mathrm{n})}(f,g)\|_{\ell^\infty(\mathbb{Z})} \lesssim_P N^{-1}\|f\|_{\ell^2(\mathbb{Z})}\|g\|_{\ell^2(\mathbb{Z})}, \tag{3.12}$$

and by interpolating this with (1.6) it is not difficult to establish (3.10) for any $1 < p_1, p_2, p \leq \infty$ with $\frac{1}{p_1} + \frac{1}{p_2} > \frac{1}{p}$. However, in this regime the Calderón transference principle no longer applies, and so no consequences to general measure preserving systems (in particular, those of finite measure) can be concluded. Indeed, the decay in N exhibited by (3.12) is not possible in the finite measure setting since $A^{\mathrm{n},P(\mathrm{n})}(1,1) = 1$. In the opposite sub-Hölder regime $\frac{1}{p_1} + \frac{1}{p_2} < \frac{1}{p}$, even single-scale boundedness $\|\tilde{A}_N^{\mathrm{n},P(\mathrm{n})}\|_{\ell^{p_1}(\mathbb{Z})\times\ell^{p_2}(\mathbb{Z})\to\ell^p(\mathbb{Z})} < \infty$ fails on the integer shift model, as can be seen by testing the operator on indicator functions of large intervals. (However, on finite measure systems one can of course deduce sub-Hölder exponent estimates from Hölder exponent estimates by applying Hölder's inequality.)

4. Abstract harmonic analysis: relating the integers to the adelic integers

We will be performing Fourier analysis on many different groups in this paper and, in particular, exploiting the close relationship between major arc Fourier analysis on the integers $\mathbb{Z}$ on the one hand, and low frequency Fourier analysis on the adelic integers $\mathbb{A}_{\mathbb{Z}}$ on the other hand (see Figure 6). It will be convenient to set out some abstract harmonic analysis notation to perform this analysis in a unified fashion. We let $\mathbb{T} := \mathbb{R}/\mathbb{Z}$ denote the unit circle, and $e: \mathbb{T} \to \mathbb{C}$ denote the standard character $e(\theta) := e^{2\pi i\theta}$.

Definition 4.1 (Pontryagin duality). An *LCA group* is a locally compact abelian group $\mathbb{G} = (\mathbb{G}, +)$ equipped with a Haar measure $\mu_{\mathbb{G}}$. A *Pontryagin dual* of an LCA group $\mathbb{G}$ is an LCA group $\mathbb{G}^* = (\mathbb{G}^*, +)$ with a Haar measure $\mu_{\mathbb{G}^*}$ and a continuous bihomomorphism $(x, \xi) \mapsto x \cdot \xi$ (which we call a *pairing*) from $\mathbb{G} \times \mathbb{G}^*$ to the unit circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, such that the Fourier transform $\mathcal{F}_{\mathbb{G}}: L^1(\mathbb{G}) \to C(\mathbb{G}^*)$ defined by

$$\mathcal{F}_{\mathbb{G}} f(\xi) := \int_{\mathbb{G}} f(x) e(x \cdot \xi)\, d\mu_{\mathbb{G}}(x)$$

extends to a unitary map from $L^2(\mathbb{G})$ to $L^2(\mathbb{G}^*)$; in particular, we have the Plancherel identity

$$\int_{\mathbb{G}} |f(x)|^2\, d\mu_{\mathbb{G}}(x) = \int_{\mathbb{G}^*} |\mathcal{F}_{\mathbb{G}} f(\xi)|^2\, d\mu_{\mathbb{G}^*}(\xi)$$

for all $f \in L^2(\mathbb{G})$.

If $\Omega \subset \mathbb{G}^*$ is measurable, we say that $f \in L^2(\mathbb{G})$ is *Fourier supported* in Ω if $\mathcal{F}_{\mathbb{G}} f$ vanishes outside of Ω (modulo null sets). The space of such functions will be denoted $L^2(\mathbb{G})^{\Omega}$.

As is well known (see, e.g., [78]), every LCA group $\mathbb{G}$ has a Pontryagin dual $\mathbb{G}^*$, and the inverse Fourier transform $\mathcal{F}_{\mathbb{G}}^{-1}: L^2(\mathbb{G}^*) \to L^2(\mathbb{G})$ is then given for $F \in L^1(\mathbb{G}^*) \cap L^2(\mathbb{G}^*)$ by the formula

$$\mathcal{F}_{\mathbb{G}}^{-1} F(x) = \int_{\mathbb{G}^*} F(\xi) e(-x \cdot \xi)\, d\mu_{\mathbb{G}^*}(\xi).$$

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

(i) If $\mathbb{G} = \mathbb{R}$ with Lebesgue measure $\mu_{\mathbb{R}} = dx$, then $\mathbb{G}^* = \mathbb{R}^* = \mathbb{R}$ with Lebesgue measure $\mu_{\mathbb{R}^*} = d\xi$ is a Pontryagin dual, with pairing $x \cdot \xi := x\xi \bmod 1$.

(ii) If $\mathbb{G} = \mathbb{Z}$ with counting measure $\mu_{\mathbb{Z}}$, then $\mathbb{G}^* = \mathbb{T}$ with Lebesgue measure $\mu_{\mathbb{T}} = d\xi$ is a Pontryagin dual, with pairing $x \cdot \xi := x\xi$.

(iii) If $\mathbb{G} = \mathbb{Z}/Q\mathbb{Z}$ is a cyclic group for some $Q \in \mathbb{Z}_+$ with normalized counting measure $\int_{\mathbb{Z}/Q\mathbb{Z}} f(x)\, d\mu_{\mathbb{Z}/Q\mathbb{Z}}(x) := \mathbb{E}_{x \in \mathbb{Z}/Q\mathbb{Z}} f(x)$, then the *dual cyclic group* $\mathbb{G}^* = \frac{1}{Q}\mathbb{Z}/\mathbb{Z}$ with counting measure $\mu_{\frac{1}{Q}\mathbb{Z}/\mathbb{Z}}$ is a Pontryagin dual, with pairing $x \cdot \xi := x\xi$.

(iv) If $\mathbb{G} = \mathbb{Z}_p := \varprojlim_j \mathbb{Z}/p^j\mathbb{Z}$ is the compact group of p-adic integers with Haar probability measure $\mu_{\mathbb{Z}_p}$ (the inverse limit of normalized counting measures on $\mathbb{Z}/p^j\mathbb{Z}$) for some prime $p \in \mathbb{P}$, then the discrete group $\mathbb{G}^* = \mathbb{Z}_p^* = \varinjlim_j \frac{1}{p^j}\mathbb{Z}/\mathbb{Z} = \mathbb{Z}[\frac{1}{p}]/\mathbb{Z}$ with counting measure $\mu_{\mathbb{Z}_p^*}$ is a Pontragin dual, with pairing $x \cdot (\frac{a}{p^j} \bmod 1) := \frac{xa \bmod p^j}{p^j}$.

(v) If $\mathbb{G} = \hat{\mathbb{Z}} := \prod_{p \in \mathbb{P}} \mathbb{Z}_p$ is the compact group of profinite integers with Haar probability measure, then the discrete group $\mathbb{G}^* = \hat{\mathbb{Z}}^* = \coprod_{p \in \mathbb{P}} \mathbb{Z}_p^* = \mathbb{Q}/\mathbb{Z}$ of "arithmetic frequencies" with counting measure $\mu_{\mathbb{Q}/\mathbb{Z}}$ is a Pontragin dual, with pairing $x \cdot (\frac{a}{q} \bmod 1) := \frac{xa \bmod q}{q}$.

(vi) If $\mathbb{G}_1, \mathbb{G}_2$ are LCA groups with Pontryagin duals $\mathbb{G}_1^*, \mathbb{G}_2^*$, then the product $\mathbb{G}_1 \times \mathbb{G}_2$ (with product Haar measure) is an LCA group with Pontryagin dual $\mathbb{G}_1^* \times \mathbb{G}_2^*$ and pairing $(x_1, x_2) \cdot (\xi_1, \xi_2) := x_1 \cdot \xi_1 + x_2 \cdot \xi_2$. In particular, if $\mathbb{G} = \mathbb{A}_{\mathbb{Z}} := \mathbb{R} \times \hat{\mathbb{Z}}$ is the adelic integers[6] (with the product Haar measure $\mu_{\mathbb{A}_{\mathbb{Z}}} := \mu_{\mathbb{R}} \times \mu_{\hat{\mathbb{Z}}}$), then *adelic frequency space* $\mathbb{G}^* = \mathbb{A}_{\mathbb{Z}}^* = \mathbb{R} \times \mathbb{Q}/\mathbb{Z}$ is a Pontryagin dual (with product measure $\mu_{\mathbb{R} \times \mathbb{Q}/\mathbb{Z}} := \mu_{\mathbb{R}} \times \mu_{\mathbb{Q}/\mathbb{Z}}$ and the indicated pairing). Similarly, for any $Q \in \mathbb{Z}_+$, $\mathbb{R} \times \mathbb{Z}/Q\mathbb{Z}$ has $\mathbb{R} \times \frac{1}{Q}\mathbb{Z}/\mathbb{Z}$ as its Pontryagin dual.

Remark 4.2. Heuristically, one can think of analysis on the adelic integers $\mathbb{A}_{\mathbb{Z}}$ (resp. the profinite integers $\hat{\mathbb{Z}}$, or the p-adic integers $\mathbb{Z}_p$) as an abstraction of analysis on the product groups $\mathbb{R} \times \mathbb{Z}/Q\mathbb{Z}$ (resp. the cyclic groups $\mathbb{Z}/Q\mathbb{Z}$, $\mathbb{Z}/p^j\mathbb{Z}$) in which all estimates are required to be uniform in the parameter Q or p^j. These abstractions are convenient to use in settings in which one does not wish to fix an ambient modulus Q or p^j in advance.

Observe that we have quotient homomorphisms $x \mapsto x \bmod Q$ from $\mathbb{Z}$ to $\mathbb{Z}/Q\mathbb{Z}$ or from $\hat{\mathbb{Z}}$ to $\mathbb{Z}/Q\mathbb{Z}$, $x \mapsto x \bmod p^j$ from $\mathbb{Z}_p$ to $\mathbb{Z}/p^j\mathbb{Z}$, and $x \mapsto x \bmod 1$ from $\mathbb{R}$ to $\mathbb{T}$. The adelic integers $\mathbb{A}_{\mathbb{Z}}$ capture two important limiting behaviors of the integers $\mathbb{Z}$: the continuous behavior (as described by the $\mathbb{R}$ factor), and the arithmetic behavior (as described by the $\hat{\mathbb{Z}}$ factor). We also have the inclusion homomorphism $\iota\colon \mathbb{Z} \to \mathbb{A}_{\mathbb{Z}}$ defined by

$$\iota(x) := \big(x, ((x \bmod p^j)_{j \in \mathbb{N}})_{p \in \mathbb{P}}\big)$$

[6]The adelic integers $\mathbb{A}_{\mathbb{Z}}$ should not be confused with the larger ring $\mathbb{A}_{\mathbb{Q}} = \mathbb{A}_{\mathbb{Z}} \otimes_{\mathbb{Z}} \mathbb{Q}$ of adelic numbers, which we will not use in this paper.

and the addition homomorphism $\pi\colon \mathbb{R}\times\mathbb{Q}/\mathbb{Z}\to\mathbb{T}$ defined by

$$\pi(\theta,\alpha) := \alpha+\theta;$$

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

$$\iota(x)\cdot\xi = x\cdot\pi(\xi) \tag{4.3}$$

for all $x\in\mathbb{Z}$ and $\xi\in\mathbb{R}\times\mathbb{Q}/\mathbb{Z}$. In "major arc" regimes we will be able to 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 $\hat{\mathbb{Z}}$. We summarized the relations between the various LCA groups in Figures 1 and 2.

Remark 4.4. As is well known, the embedding ι identifies $\mathbb{Z}$ with a cocompact lattice $\iota(\mathbb{Z})$ in $\mathbb{A}_{\mathbb{Z}}$. (Thus $\iota(\mathbb{Z})$ is a discrete subgroup of $\mathbb{A}_{\mathbb{Z}}$ and the quotient $\mathbb{A}_{\mathbb{Z}}/\iota(\mathbb{Z})$ is compact.) Thus $\mathbb{A}_{\mathbb{Z}}$ is in some sense only "slightly" larger than $\mathbb{Z}$ itself, but has the advantage of splitting completely into a continuous component $\mathbb{R}$ and an arithmetic component $\hat{\mathbb{Z}}$, whereas $\mathbb{Z}$ does not directly have such a splitting. However, the point is that after restricting attention to major arcs, one can partially move back and forth between the integers and adelic integers, and thus have some chance of exploiting the product structure of $\mathbb{A}_{\mathbb{Z}} = \mathbb{R}\times\hat{\mathbb{Z}}$ to decouple the continuous and arithmetic aspects of the analysis.

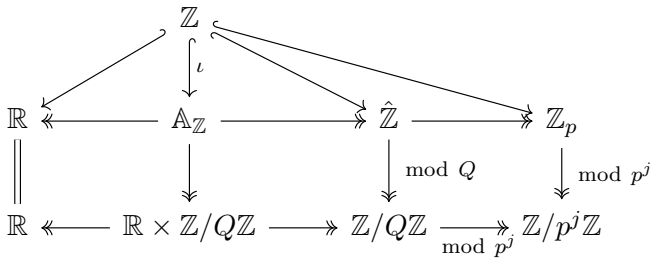


Figure 1. *A commutative diagram of the various physical space LCA groups used in this paper, with the arrows indicating continuous homomorphisms.* Here Q is a positive integer, and p^j is a prime power dividing Q. Double-headed arrows are surjective; arrows with hooks are injective. The left column contains "continuous" groups, the right two columns contain "arithmetic" groups (and are compact), and the second column from the left contain groups exhibiting both continuous and arithmetic aspects. The second row is the inverse limit of the third. Note the central role played by the adelic integers $\mathbb{A}_{\mathbb{Z}}$.

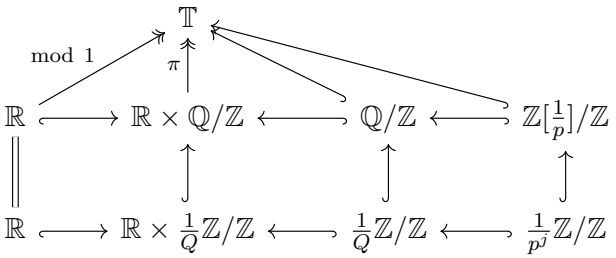


Figure 2. *A commutative diagram of the various frequency space LCA groups used in this paper. The groups in the right two columns are discrete.* The second row is the direct limit of the third. Note the duality with Figure 1. (This can be made precise using Fourier adjoint relationships such as (4.3).)

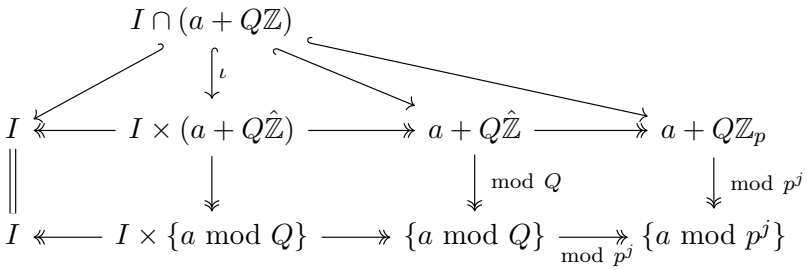


Figure 3. *A restriction of the physical space diagram in Figure 1 to an arithmetic progression $I \cap (a+Q\mathbb{Z})$ formed by intersecting an interval $I \subset \mathbb{R}$ with an infinite arithmetic progression $a + Q\mathbb{Z}$.* The sets here are no longer groups in general (except in an "approximate" sense), and so the arrows no longer denote homomorphisms. As in the previous figures, p^j is understood to be a prime power dividing Q. Note how this diagram separates an arithmetic progression into its continuous and arithmetic components.

For various LCA groups $\mathbb{G}$, we shall work with a space $\boldsymbol{S}(\mathbb{G}) \subset L^1(\mathbb{G}) \cap L^\infty(\mathbb{G})$ of Schwartz–Bruhat functions $f\colon G \to \mathbb{C}$, generalizing the classical class of Schwartz functions on $\mathbb{R}$ that serve as a useful class of "nice" functions that are dense in $L^p(\mathbb{G})$ for every $1 \le p < \infty$ and behave well with respect to Fourier-analytic operations. A definition of this space for arbitrary LCA groups can be found, for instance, in [16], [71], but for the purpose of this paper we shall only need the following special cases:

(i) $\boldsymbol{S}(\mathbb{R})$ is the space of Schwartz functions on $\mathbb{R}$.
(ii) $\boldsymbol{S}(\mathbb{Z})$ is the space of rapidly decreasing functions on $\mathbb{Z}$, and $\boldsymbol{S}(\mathbb{T})$ is the space of smooth functions on $\mathbb{T}$.

(iii) $\boldsymbol{S}(\mathbb{Z}/Q\mathbb{Z})$ is the space of arbitrary functions on $\mathbb{Z}/Q\mathbb{Z}$, and similarly for $\boldsymbol{S}(\frac{1}{Q}\mathbb{Z}/\mathbb{Z})$.

(iv) $\boldsymbol{S}(\mathbb{Z}_p)$ is the space of locally constant functions f on $\mathbb{Z}_p$, or equivalently those functions of the form $f(x) = f_j(x \bmod p^j)$ for some $j \in \mathbb{N}$ and some function $f_j\colon \mathbb{Z}/p^j\mathbb{Z} \to \mathbb{C}$. $\boldsymbol{S}(\mathbb{Z}_p^*)$ is the space of finitely supported functions on $\mathbb{Z}_p^*$.

(v) $\boldsymbol{S}(\hat{\mathbb{Z}})$ is the space of locally constant functions f on $\hat{\mathbb{Z}}$, or equivalently those functions of the form $f(x) = f_Q(x \bmod Q)$ for some $Q \in \mathbb{Z}_+$ and $f_Q\colon \mathbb{Z}/Q\mathbb{Z} \to \mathbb{C}$. $\boldsymbol{S}(\hat{\mathbb{Z}}^*)$ is the space of finitely supported functions on $\hat{\mathbb{Z}}^*$.

(vi) $\boldsymbol{S}(\mathbb{R} \times \mathbb{Z}/Q\mathbb{Z})$ is the space of functions that is Schwartz in the $\mathbb{R}$ variable, and similarly for $\boldsymbol{S}(\mathbb{R} \times \frac{1}{Q}\mathbb{Z}/\mathbb{Z})$.

(vii) $\boldsymbol{S}(\mathbb{A}_{\mathbb{Z}})$ is the space of functions of the form $f(x,y) = f_Q(x, y \bmod Q)$ for some $Q \in \mathbb{Z}_+$ and $f_Q\colon \mathbb{R} \times \mathbb{Z}/Q\mathbb{Z} \to \mathbb{C}$ that is Schwartz in the $\mathbb{R}$ variable. $\boldsymbol{S}(\mathbb{R} \times \mathbb{Q}/\mathbb{Z})$ is the space of functions supported on $\mathbb{R} \times \Sigma$ for some finite set $\Sigma \subset \mathbb{Q}/\mathbb{Z}$ and Schwartz in the $\mathbb{R}$ variable.

(viii) If $\mathbb{G}_1, \mathbb{G}_2$ are any two of the groups listed above, we define the Schwartz–Bruhat space $\boldsymbol{S}(\mathbb{G}_1 \times \mathbb{G}_2)$ on the product LCA group $\mathbb{G}_1 \times \mathbb{G}_2$ in the obvious fashion. We note that if $f_1 \in \boldsymbol{S}(\mathbb{G}_1)$ and $f_2 \in \boldsymbol{S}(\mathbb{G}_2)$, then $f_1 \otimes f_2$ can be identified with an element of $\boldsymbol{S}(\mathbb{G}_1 \times \mathbb{G}_2)$.

One could place a topology on the Schwartz–Bruhat spaces $\boldsymbol{S}(\mathbb{G})$, but we will not need to do so here. As is well known, the Fourier transform $\mathcal{F}_{\mathbb{G}}$ is a bijection from $\boldsymbol{S}(\mathbb{G})$ to $\boldsymbol{S}(\mathbb{G}^*)$ for any of the groups $\mathbb{G}$ in Figure 1. The Fourier transform can also be extended to vector-valued functions taking values in a finite-dimensional vector space V in the obvious fashion.

If $\Omega \subset \mathbb{G}^*$, we let $\boldsymbol{S}(\mathbb{G})^{\Omega}$ denote the subspace of $\boldsymbol{S}(\mathbb{G})$ consisting of functions that are Fourier supported on Ω, and we let $\boldsymbol{S}(\Omega)$ denote the subspace of $\boldsymbol{S}(\mathbb{G}^*)$ consisting of functions that are supported on Ω. Thus $\mathcal{F}_{\mathbb{G}}$ is also a bijection between $\boldsymbol{S}(\mathbb{G})^{\Omega}$ and $\boldsymbol{S}(\Omega)$.

The inclusion homomorphism $\iota\colon \mathbb{Z} \to \mathbb{A}_{\mathbb{Z}}$ gives rise to a *sampling map* $\mathcal{S}\colon \boldsymbol{S}(\mathbb{A}_{\mathbb{Z}}) \to \boldsymbol{S}(\mathbb{Z})$ defined by

$$\mathcal{S}f(x) \coloneqq f(\iota(x))$$

for $x \in \mathbb{Z}$ and $f \in \boldsymbol{S}(\mathbb{A}_{\mathbb{Z}})$. Dually, the addition homomorphism $\pi\colon \mathbb{R}\times\mathbb{Q}/\mathbb{Z} \to \mathbb{T}$ gives rise to a *projection map* $\mathcal{P}\colon \boldsymbol{S}(\mathbb{R} \times \mathbb{Q}/\mathbb{Z}) \to \boldsymbol{S}(\mathbb{T})$, defined by the formula

$$\mathcal{P}F(\xi) \coloneqq \sum_{(\theta,\alpha)\in\pi^{-1}(\xi)} F(\theta,\alpha)$$

for $\theta \in \mathbb{R}$, $\alpha \in \mathbb{Q}/\mathbb{Z}$, and $F \in \boldsymbol{S}(\mathbb{R} \times \mathbb{Q}/\mathbb{Z})$. (Note that the definition of $\boldsymbol{S}(\mathbb{R} \times \mathbb{Q}/\mathbb{Z})$ ensures that this sum contains at most countably many non-zero terms.) From (4.3), one has the identity

$$\mathcal{F}_{\mathbb{Z}}^{-1} \circ \mathcal{P} = \mathcal{S} \circ \mathcal{F}_{\mathbb{A}_{\mathbb{Z}}}^{-1}$$

or equivalently the adelic Poisson summation formula

$$\mathcal{F}_{\mathbb{Z}} \circ \mathcal{S} = \mathcal{P} \circ \mathcal{F}_{\mathbb{A}_{\mathbb{Z}}},$$

and so we have the commutative diagram

$$\begin{array}{ccc} \boldsymbol{S}(\mathbb{Z}) & \xleftarrow[\mathcal{S}]{} & \boldsymbol{S}(\mathbb{A}_{\mathbb{Z}}) \\ \downarrow{\scriptstyle \mathcal{F}_{\mathbb{Z}}} & & \downarrow{\scriptstyle \mathcal{F}_{\mathbb{A}_{\mathbb{Z}}}} \\ \boldsymbol{S}(\mathbb{T}) & \xleftarrow[\mathcal{P}]{} & \boldsymbol{S}(\mathbb{R} \times \mathbb{Q}/\mathbb{Z}); \end{array}$$

see also Figures 4 and 5.

A key difficulty here is that of *aliasing*: the non-injectivity of $\pi\colon \mathbb{R} \times \mathbb{Q}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ causes the sampling map $\mathcal{S}\colon \boldsymbol{S}(\mathbb{A}_{\mathbb{Z}}) \to \boldsymbol{S}(\mathbb{Z})$ to also be non-injective. Indeed, if $(\xi_1, \alpha_1), (\xi_2, \alpha_2)$ are distinct elements of $\mathbb{R} \times \mathbb{Q}/\mathbb{Z}$ and are such that $\pi(\xi_1, \alpha_1) = \pi(\xi_2, \alpha_2)$, then for any non-zero $F \in \boldsymbol{S}(\mathbb{A}_{\mathbb{Z}})$, the functions $F_1(x, y) := e(x\xi_1 + y\cdot\alpha_1)F(x, y)$ and $F_2(x, y) := e(x\xi_2 + y\cdot\alpha_2)F(x, y)$ are distinct elements of $\boldsymbol{S}(\mathbb{A}_{\mathbb{Z}})$ that are "aliased" in the sense that $\mathcal{S}F_1 = \mathcal{S}F_2$. However, we can avoid this problem by restricting attention to a compact subset Ω of adelic frequency space $\mathbb{R} \times \mathbb{Q}/\mathbb{Z}$ that is *non-aliasing* in the sense that the addition homomorphism π is injective on Ω, so that $\mathcal{P}$ becomes an algebra homomorphism from $\boldsymbol{S}(\Omega)$ to $\boldsymbol{S}(\pi(\Omega))$. Thus

$$\mathcal{P}(FG) = \mathcal{P}(F)\mathcal{P}(G) \tag{4.5}$$

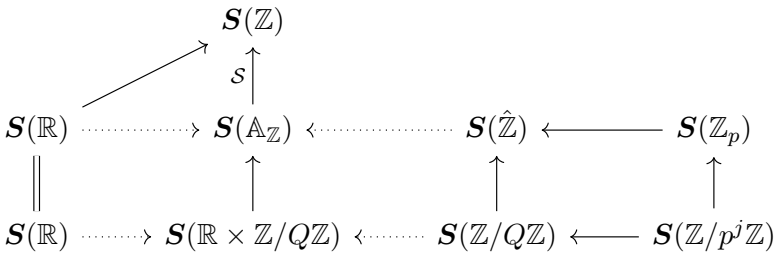


Figure 4. *Schwartz–Bruhat spaces on physical space LCA groups.* Solid arrows indicate canonical linear maps of a "sampling" or "pullback" nature; dotted arrows from two spaces V_1, V_2 to a third V indicate the existence of a tensor product operation $\otimes\colon V_1 \times V_2 \to V$. The second row is the direct limit of the third. Compare with Figure 1. (Some arrows in that figure do not have an analogue here, basically because $\boldsymbol{S}(\mathbb{R})$ does not contain a multiplicative unit 1, and the inclusions of $\mathbb{Z}$ into $\hat{\mathbb{Z}}$ and $\mathbb{Z}_p$ are not proper.)

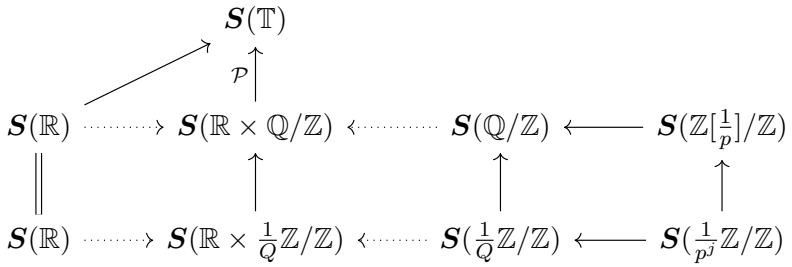


Figure 5. *Schwartz–Bruhat spaces on frequency space LCA groups.* Solid arrows indicate canonical linear maps of a "projection" or "pushforward" nature; dotted arrows indicate a tensor product as in Figure 4. The second row is the direct limit of the third. This figure and the preceding one are intertwined by the Fourier transform via various forms of the Poisson summation formula. Compare also with Figure 2. (Some arrows in that figure do not have an analogue here, basically because $\boldsymbol{S}(\mathbb{R})$ does not contain a convolution unit δ, and the embeddings of $\mathbb{Q}/\mathbb{Z}$ and $\mathbb{Z}[\frac{1}{p}]/\mathbb{Z}$ into $\mathbb{T}$ are not open.)

for all $F, G \in \boldsymbol{S}(\Omega)$, and one has the commutative diagram

$$\begin{array}{ccc} \boldsymbol{S}(\mathbb{Z})^{\pi(\Omega)} & \xleftarrow[\mathcal{S}]{} & \boldsymbol{S}(\mathbb{A}_{\mathbb{Z}})^{\Omega} \\ \downarrow{\scriptstyle \mathcal{F}_{\mathbb{Z}}} & & \downarrow{\scriptstyle \mathcal{F}_{\mathbb{A}_{\mathbb{Z}}}} \\ \boldsymbol{S}(\pi(\Omega)) & \xleftarrow[\mathcal{P}]{} & \boldsymbol{S}(\Omega) \end{array} . \tag{4.6}$$

In this case one verifies that the lower three maps $\mathcal{F}_{\mathbb{Z}}, \mathcal{P}, \mathcal{F}_{\mathbb{A}_{\mathbb{Z}}}$ are invertible, hence the upper map $\mathcal{S}$ is also. In particular, to any non-aliasing compact set of adelic frequencies Ω, we can associate an *interpolation operator* $\mathcal{S}_{\Omega}^{-1}\colon \boldsymbol{S}(\mathbb{Z})^{\pi(\Omega)} \to \boldsymbol{S}(\mathbb{A}_{\mathbb{Z}})^{\Omega}$ that extends any Schwartz–Bruhat function on the integers with Fourier support in $\pi(\Omega)$ to the unique Schwartz–Bruhat extension on the adelic integers with Fourier support in Ω. Note from (4.6) and Plancherel's theorem that the sampling operator $\mathcal{S}$ and the interpolation operator $\mathcal{S}_{\Omega}^{-1}$ extend to unitary maps between $\ell^2(\mathbb{Z})^{\pi(\Omega)}$ and $L^2(\mathbb{A}_{\mathbb{Z}})^{\Omega}$ that invert each other.

The diagram (4.6) allows us to equate certain portions of Fourier analysis on the integers $\mathbb{Z}$ with corresponding portions of Fourier analysis of the adelic integers $\mathbb{A}_{\mathbb{Z}}$; this will be useful for clarifying Fourier analysis on major arcs $\mathcal{M}_{\leq l,\leq k}$, which in this perspective are interpreted as projections of a certain non-aliasing Cartesian product $\mathbb{R}_{\leq l} \times (\mathbb{Q}/\mathbb{Z})_{\leq k}$ of adelic frequency space; see Figure 6 and Section 5 for definitions.

Example 4.7. If $Q \in \mathbb{Z}_+$ and $r > 0$, then $[-r, r] \times \frac{1}{Q}\mathbb{Z}/\mathbb{Z}$ is non-aliasing if and only if $r < \frac{1}{2Q}$. The injectivity of $\mathcal{S}$ in this case is a variant of the classical Shannon sampling theorem. See also Theorem 4.18 below.

Now we define Fourier multiplier operators. A continuous function $\varphi\colon \mathbb{G}^* \to \mathbb{C}$ is said to be *smooth tempered* if $\varphi F \in \boldsymbol{S}(\mathbb{G}^*)$ whenever $F \in \boldsymbol{S}(\mathbb{G}^*)$. For instance, $\varphi\colon \mathbb{R} \to \mathbb{C}$ is smooth tempered if and only if all derivatives exist and are of at most polynomial growth.

Definition 4.8 (Fourier multiplier operators). Let $\mathbb{G}$ be one of the LCA groups in Figure 1.

(i) If $\varphi\colon \mathbb{G}^* \to \mathbb{C}$ is a smooth tempered function, we define the Fourier multiplier operator $\mathrm{T}_\varphi\colon \boldsymbol{S}(\mathbb{G}) \to \boldsymbol{S}(\mathbb{G})$ by the formula

$$\mathcal{F}_{\mathbb{G}} \mathrm{T}_\varphi = \varphi \mathcal{F}_{\mathbb{G}}$$

or equivalently

$$\mathrm{T}_\varphi f(x) = \int_{\mathbb{G}^*} \varphi(\xi) \mathcal{F}_{\mathbb{G}} f(\xi) e(-x \cdot \xi)\, d\mu_{\mathbb{G}^*}(\xi)$$

for $f \in \boldsymbol{S}(\mathbb{G})$ and $x \in \mathbb{G}$. We refer to φ as the *symbol* of T_φ.

(ii) If $m\colon \mathbb{G}^* \times \mathbb{G}^* \to \mathbb{C}$ is a smooth tempered function, we define the bilinear Fourier multiplier operator $\mathrm{B}_m\colon \boldsymbol{S}(\mathbb{G}) \times \boldsymbol{S}(\mathbb{G}) \to \boldsymbol{S}(\mathbb{G})$ by the formula

$$\mathrm{B}_m(f,g)(x) = \int_{\mathbb{G}^*} \int_{\mathbb{G}^*} m(\xi_1, \xi_2) \mathcal{F}_{\mathbb{G}} f(\xi_1) \mathcal{F}_{\mathbb{G}} g(\xi_2) e(-x \cdot (\xi_1 + \xi_2))\, d\mu_{\mathbb{G}^*}(\xi_1) d\mu_{\mathbb{G}^*}(\xi_2).$$

We refer to m as the *symbol* of B_m.

Clearly T_φ depends linearly on φ, and B_m depends linearly on m. We also observe the functional calculus identities

$$\begin{aligned} \mathrm{T}_1 f &= f, \\ \mathrm{B}_1(f,g) &= fg, \\ \mathrm{T}_{\varphi_1\varphi_2} f &= \mathrm{T}_{\varphi_1} \mathrm{T}_{\varphi_2} f, \\ \mathrm{B}_{m(\varphi_1 \otimes \varphi_2)}(f,g) &= \mathrm{B}_m(\mathrm{T}_{\varphi_1} f, \mathrm{T}_{\varphi_2} g) \end{aligned} \tag{4.9}$$

whenever $f, g \in \boldsymbol{S}(\mathbb{G})$ and φ_1, φ_2, m are smooth tempered functions on $\mathbb{G}^*$, $\mathbb{G}^*$, $\mathbb{G}^* \times \mathbb{G}^*$ respectively. Finally we observe that T_φ is self-adjoint on $L^2(\mathbb{G})$ when φ is real-valued. We can also extend the linear Fourier multipliers T_φ to Schwartz–Bruhat functions $\boldsymbol{S}(\mathbb{G}; V)$ taking values in a finite-dimensional vector space V in the obvious fashion.

Example 4.10 (Averaging operators as Fourier multipliers). We work on the integer shift system. If $P \in \mathbb{Z}[\mathrm{n}]$, the averaging operator $A_N^{P(\mathrm{n})}$ is a linear

Fourier multiplier operator on $\boldsymbol{S}(\mathbb{Z})$ with symbol

$$\varphi_{N,\mathbb{Z}}(\xi) \coloneqq \mathbb{E}_{n\in[N]} e(P(n)\xi)$$

for $\xi \in \mathbb{T}$. Similarly, if $P_1, P_2 \in \mathbb{Z}[\mathrm{n}]$, then the averaging operator $A_N^{P_1(\mathrm{n}),P_2(\mathrm{n})}$ is a bilinear Fourier multiplier operator on $\boldsymbol{S}(\mathbb{Z})$ with symbol

$$m_{N,\mathbb{Z}}(\xi_1,\xi_2) \coloneqq \mathbb{E}_{n\in[N]} e(P_1(n)\xi_1 + P_2(n)\xi_2)$$

and $\tilde{A}_N^{P_1(\mathrm{n}),P_2(\mathrm{n})}$ similarly has symbol

$$\tilde{m}_{N,\mathbb{Z}}(\xi_1,\xi_2) \coloneqq \mathbb{E}_{n\in[N]} e(P_1(n)\xi_1 + P_2(n)\xi_2)\mathbb{1}_{n>N/2}$$

for $\xi_1, \xi_2 \in \mathbb{T}$. If $\mathbb{G}$ is one of the compact rings $\mathbb{Z}/p^j\mathbb{Z}$, $\mathbb{Z}/Q\mathbb{Z}$, $\mathbb{Z}_p$, or $\hat{\mathbb{Z}}$, then P_1, P_2 can be thought of as continuous maps from $\mathbb{G}$ to itself, and we can define the averaging operator $A_{\mathbb{G}} = A_{\mathbb{G}}^{P_1(\mathrm{n}),P_2(\mathrm{n})} \colon \boldsymbol{S}(\mathbb{G}) \times \boldsymbol{S}(\mathbb{G}) \to \boldsymbol{S}(\mathbb{G})$ by the formula

$$A_{\mathbb{G}}(f,g)(x) \coloneqq \int_{\mathbb{G}} f(x - P_1(y))g(x - P_2(y))\, d\mu_{\mathbb{G}}(y).$$

From the Fourier inversion formula and the Fubini–Tonelli theorem, we see that $A_{\mathbb{G}}$ is a bilinear Fourier multiplier operator with symbol

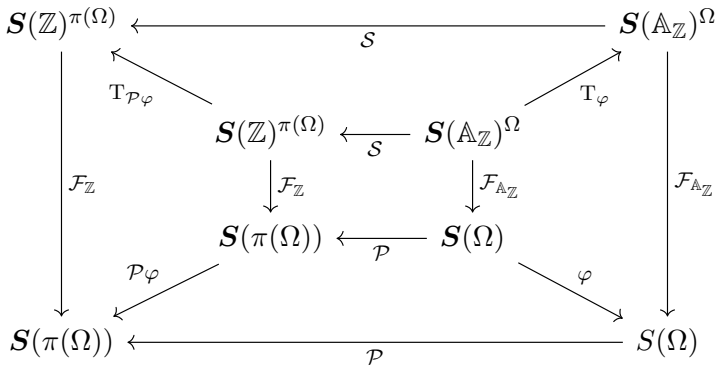
$$m_{\mathbb{G}}(\xi_1,\xi_2) \coloneqq \int_{\mathbb{G}} e(P_1(y)\xi_1 + P_2(y)\xi_2)\, d\mu_{\mathbb{G}}(y)$$

for $\xi_1, \xi_2 \in \mathbb{G}^*$.

Example 4.11 (Tensor products of multipliers). Let $\mathbb{G}_1, \mathbb{G}_2$ be LCA groups from Figure 1. If T_{φ_1} is a linear Fourier multiplier operator on $\boldsymbol{S}(\mathbb{G}_1)$ and T_{φ_2} is a linear Fourier multiplier operator on $\boldsymbol{S}(\mathbb{G}_2)$, then $\mathrm{T}_{\varphi_1\otimes\varphi_2}$ is a linear Fourier multiplier operator on $\boldsymbol{S}(\mathbb{G}_1 \times \mathbb{G}_2)$ that is the tensor product of T_{φ_1} and T_{φ_2} in the sense that (2.9) holds for all $f_1 \in \boldsymbol{S}(\mathbb{G}_1)$, $f_2 \in \boldsymbol{S}(\mathbb{G}_2)$. Similarly, if $\mathrm{B}_{m_1}, \mathrm{B}_{m_2}$ are bilinear Fourier multiplier operators on $\boldsymbol{S}(\mathbb{G}_1), \boldsymbol{S}(\mathbb{G}_2)$ respectively, then the bilinear Fourier multiplier operator $\mathrm{B}_{m_1\otimes m_2}$ is the tensor product of B_{m_1} and B_{m_2} in the sense that (2.10) holds for all $f_1, g_1 \in \boldsymbol{S}(\mathbb{G}_1)$, $f_2, g_2 \in \boldsymbol{S}(\mathbb{G}_2)$.

As previously mentioned, if Ω is a non-aliasing subset of $\mathbb{R} \times \mathbb{Q}/\mathbb{Z}$, then the sampling operator $\mathcal{S}$ restricts to a unitary map from $L^2(\mathbb{A}_{\mathbb{Z}})^{\Omega}$ to $\ell^2(\mathbb{Z})^{\pi(\Omega)}$, or equivalently the interpolation operator $\mathcal{S}_{\Omega}^{-1}$ is a unitary map from $\ell^2(\mathbb{Z})^{\pi(\Omega)}$ to $L^2(\mathbb{A}_{\mathbb{Z}})^{\Omega}$. This suggests that Fourier multiplier operators on $L^2(\mathbb{A}_{\mathbb{Z}})^{\Omega}$ can be identified with Fourier multiplier operators on $\ell^2(\mathbb{Z})^{\pi(\Omega)}$. This is indeed the case:

Lemma 4.12 (Adelic and integer Fourier multipliers). *Let $\Omega \subset \mathbb{R} \times \mathbb{Q}/\mathbb{Z}$ be a non-aliasing compact subset of adelic frequency space. Then for any $\varphi \in \boldsymbol{S}(\Omega)$, the following diagram commutes, where φ denotes the operation of pointwise multiplication by φ, and similarly for $\mathcal{P}\varphi$:*



In particular, one has

$$\mathrm{T}_{\mathcal{P}\varphi}\mathcal{S}f = \mathcal{S}\mathrm{T}_{\varphi}f \tag{4.13}$$

for all $f \in \boldsymbol{S}(\mathbb{A}_{\mathbb{Z}})^{\Omega}$.

Proof. This is immediate from (4.6), (4.5), Definition 4.8, and a routine diagram chase using the invertibility of the Fourier transform. □

Another way of writing (4.13) is as

$$\mathrm{T}_{\mathcal{P}\varphi}f = \mathcal{S}\mathrm{T}_{\varphi}\mathcal{S}_{\Omega}^{-1}f \tag{4.14}$$

for all $f \in \boldsymbol{S}(\mathbb{Z})^{\pi(\Omega)}$.

There is a bilinear version of the formula (4.13). Define the tensor square $\mathcal{P}^{\otimes 2}\colon \boldsymbol{S}((\mathbb{R}\times\mathbb{Q}/\mathbb{Z})^2) \to \boldsymbol{S}(\mathbb{T}^2)$ of the projection operator $\mathcal{P}$ by the formula

$$\mathcal{P}^{\otimes 2}m(\xi_1,\xi_2) := \sum_{(\theta_1,\alpha_1)\in\pi^{-1}(\xi_1)}\ \sum_{(\theta_2,\alpha_2)\in\pi^{-1}(\xi_2)} m((\theta_1,\alpha_1),(\theta_2,\alpha_2))$$

for all $m \in \boldsymbol{S}((\mathbb{R}\times\mathbb{Q}/\mathbb{Z})^2)$. If $\Omega_1,\Omega_2 \subset \mathbb{R}\times\mathbb{Q}/\mathbb{Z}$ are non-aliasing compact subsets of adelic frequency space, then $\mathcal{P}^{\otimes 2}$ is an algebra homomorphism from $\boldsymbol{S}(\Omega_1\times\Omega_2)$ to $\boldsymbol{S}(\pi(\Omega_1)\times\pi(\Omega_2))$, and is the tensor product of the algebra homomorphisms $\mathcal{P}\colon \boldsymbol{S}(\Omega_1) \to \boldsymbol{S}(\pi(\Omega_1))$ and $\mathcal{P}\colon \boldsymbol{S}(\Omega_2) \to \boldsymbol{S}(\pi(\Omega_2))$ in the sense of (2.9). A routine calculation (or a chase of a more complicated version of the commutative diagram in Lemma 4.12) then gives the bilinear variant

$$\mathrm{B}_{\mathcal{P}^{\otimes 2}m}(\mathcal{S}f,\mathcal{S}g) = \mathcal{S}\mathrm{B}_m(f,g) \tag{4.15}$$

of (4.13) whenever $f\in\boldsymbol{S}(\mathbb{A}_{\mathbb{Z}})^{\Omega_1}$, $g\in\boldsymbol{S}(\mathbb{A}_{\mathbb{Z}})^{\Omega_2}$, and $m\in\boldsymbol{S}(\Omega_1\times\Omega_2)$; equivalently, one has

$$\mathrm{B}_{\mathcal{P}^{\otimes 2}m}(f,g) = \mathcal{S}\mathrm{B}_m(\mathcal{S}_{\Omega_1}^{-1}f,\mathcal{S}_{\Omega_2}^{-1}g) \tag{4.16}$$

whenever $f \in \boldsymbol{S}(\mathbb{Z})^{\pi(\Omega_1)}$, $g \in \boldsymbol{S}(\mathbb{Z})^{\pi(\Omega_2)}$. From (4.9) we also observe the projected functional calculus

$$\mathrm{B}_{\mathcal{P}^{\otimes 2}m}(\mathrm{T}_{\mathcal{P}\varphi_1}f,\mathrm{T}_{\mathcal{P}\varphi_2}g) = \mathrm{B}_{\mathcal{P}^{\otimes 2}(m(\varphi_1\otimes\varphi_2))}(f,g) \tag{4.17}$$

whenever $f \in \boldsymbol{S}(\mathbb{Z})^{\pi(\Omega_1)}$, $g \in \boldsymbol{S}(\mathbb{Z})^{\pi(\Omega_2)}$, $\varphi_1 \in \boldsymbol{S}(\Omega_1)$, $\varphi_2 \in \boldsymbol{S}(\Omega_2)$, and $m \in \boldsymbol{S}(\Omega_1 \times \Omega_2)$.

The point of the identities (4.14) and (4.16) is that complicated linear and bilinear Fourier multiplier operators $\mathrm{T}_{\mathcal{P}\varphi}, \mathrm{B}_{\mathcal{P}^{\otimes 2}m}$ on the integers $\mathbb{Z}$ can be expressed (in non-aliasing regions of adelic frequency space) by simpler linear and bilinear Fourier multiplier operators $\mathrm{T}_\varphi, \mathrm{B}_m$ on the adelic integers $\mathbb{A}_\mathbb{Z}$. For the multiplier operators of interest in this paper, the adelic symbols φ, m often have a tensor product structure (or at least can be decomposed or approximated by symbols with such a structure), allowing us to decouple the Fourier analysis into the continuous Fourier analysis of $\mathbb{R}$ and the arithmetic Fourier analysis of $\hat{\mathbb{Z}}$. In many cases the arithmetic symbol factors further, allowing one to work on smaller factor groups such as $\mathbb{Z}/Q\mathbb{Z}$, $\mathbb{Z}_p$, or $\mathbb{Z}/p^j\mathbb{Z}$.

As already observed, whenever Ω is a non-aliasing compact subset of $\mathbb{R} \times \mathbb{Q}/\mathbb{Z}$, the sampling operator $\mathcal{S}\colon \boldsymbol{S}(\mathbb{A}_\mathbb{Z})^\Omega \to \boldsymbol{S}(\mathbb{Z})^{\pi(\Omega)}$ and the interpolation operator $\mathcal{S}_\Omega^{-1}\colon \boldsymbol{S}(\mathbb{Z})^{\pi(\Omega)} \to \boldsymbol{S}(\mathbb{A}_\mathbb{Z})^\Omega$ both preserve the L^2 norm. The situation for other function space norms is less clear. However, the situation is particularly favorable in the case of Example 4.7, in that the sampling and interpolation operators essentially preserve all L^p norms, even for non-Banach exponents $0 < p < 1$ or for vector-valued functions (or both):

THEOREM 4.18 (Quantitative Shannon sampling theorem). *Let $0 < p \leq \infty$, and let B be a finite-dimensional normed vector space. If $F \in \boldsymbol{S}(\mathbb{A}_\mathbb{Z}; B)$ 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 $0 < c_0 < \frac{1}{2}$, then*

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

where we extend the sampling operator $\mathcal{S}$ to vector-valued functions in the obvious fashion.

See also the sampling principle of Magyar–Stein–Wainger [64, Cor. 2.1, pp. 196] as well as [69, Prop. 4.4, pp. 816] for closely related statements. Theorem 4.18 implies that if Ω is a compact subset of $[-\frac{c_0}{Q}, \frac{c_0}{Q}] \times \frac{1}{Q}\mathbb{Z}/\mathbb{Z}$, then $\mathcal{S}\colon \boldsymbol{S}(\mathbb{A}_\mathbb{Z})^\Omega \to \boldsymbol{S}(\mathbb{Z})^{\pi(\Omega)}$ and $\mathcal{S}_\Omega^{-1}\colon \boldsymbol{S}(\mathbb{Z})^{\pi(\Omega)} \to \boldsymbol{S}(\mathbb{A}_\mathbb{Z})^\Omega$ are both bounded on L^p with norm $O_{c_0}(1)$.

Proof. As F has Fourier support on the Pontryagin dual $\mathbb{R} \times \frac{1}{Q}\mathbb{Z}/\mathbb{Z}$ of $\mathbb{R} \times \mathbb{Z}/Q\mathbb{Z}$, we can descend to the quotient group $\mathbb{R} \times \mathbb{Z}/Q\mathbb{Z}$ and establish the bound

$$\|\mathcal{S}_Q F\|_{\ell^p(\mathbb{Z};B)} \sim_{c_0} \|F\|_{L^p(\mathbb{R}\times\mathbb{Z}/Q\mathbb{Z};B)}$$

whenever $F \in \boldsymbol{S}(\mathbb{R} \times \mathbb{Z}/Q\mathbb{Z}; B)$ has Fourier support in $[-\frac{c_0}{Q}, \frac{c_0}{Q}] \times \frac{1}{Q}\mathbb{Z}/\mathbb{Z}$ and

$$\mathcal{S}_Q F(x) \coloneqq F(x, x \bmod Q).$$

By splitting $\mathbb{Z}$ into residue classes $a + Q\mathbb{Z}$ for $a \in [Q]$, and similarly splitting $\mathbb{R} \times \mathbb{Z}/Q\mathbb{Z}$ into copies $\mathbb{R} \times \{a \bmod Q\}$ of Q, it suffices by the Fubini–Tonelli

theorem to establish the bound

$$\|f\|_{\ell^p(a+Q\mathbb{Z};B)} \sim_{c_0,p} Q^{-1/p}\|f\|_{L^p(\mathbb{R};B)}$$

whenever $a \in [Q]$ and whenever $f \in \boldsymbol{S}(\mathbb{R};B)$ has Fourier support in $[-\frac{c_0}{Q},\frac{c_0}{Q}]$. After applying translation and rescaling, it suffices to show that

$$\|f\|_{\ell^p(\mathbb{Z};B)} \sim_{c_0,p} \|f\|_{L^p(\mathbb{R};B)}$$

whenever $f \in \boldsymbol{S}(\mathbb{R};B)$ has Fourier support in $[-c_0,c_0]$. It will suffice to establish the bound

$$\|f\|_{\ell^p(\mathbb{Z}+\theta;B)} \sim_{c_0,p} \|f\|_{\ell^p(\mathbb{Z};B)}$$

uniformly for all $0 \le \theta \le 1$, as the claim then follows by taking L^p norms in θ and applying the Fubini–Tonelli theorem. By translation and reflection symmetry it suffices to establish the upper bound

$$\|f\|_{\ell^p(\mathbb{Z}+\theta;B)} \lesssim_{c_0,p} \|f\|_{L^p(\mathbb{Z};B)}. \tag{4.19}$$

Let $\psi = \psi_{c_0} \in \boldsymbol{S}(\mathbb{R})$ be a function chosen so that $\mathcal{F}_{\mathbb{R}}\psi$ is supported on $[-1/2,1/2]$ and equals one on $[-c_0,c_0]$, so that the upper bound now follows from Schur's test. From the Poisson summation formula we have

$$f(y) = \sum_{x\in\mathbb{Z}} \psi(y-x)f(x)$$

for all $y \in \mathbb{R}$, hence by the triangle inequality

$$\|f(y)\|_B \le \sum_{x\in\mathbb{Z}} |\psi(y-x)|\|f(x)\|_B.$$

For $p \ge 1$, this gives (4.19) from Schur's test and the rapid decrease of ψ. For $p < 1$, we use the previous inequality to obtain

$$\|f(y)\|_B^p \le \sum_{x\in\mathbb{Z}} |\psi(y-x)|^p\|f(x)\|_B^p.$$

The claim follows from the triangle inequality and the rapid decrease of ψ. □

Because of this theorem and (4.13) and (4.15), the L^p multiplier theory for both linear and bilinear Fourier multiplier operators $\mathrm{T}_{\mathcal{P}_\varphi}$, $\mathrm{B}_{\mathcal{P}^{\otimes 2}m}$ on $\boldsymbol{S}(\mathbb{Z})^{\pi(\Omega)}$ can be easily transferred to the corresponding L^p multiplier theory of $\mathrm{T}_\varphi, \mathrm{B}_m$ on $\boldsymbol{S}(\mathbb{A}_{\mathbb{Z}})^{\Omega}$ when Ω is of the form in Example 4.7 (or a compact subset of that example). Unfortunately, this situation only occurs for us in certain "large-scale" settings, in which the widths of the major arcs are extremely narrow compared to the height. In the opposite "small-scale" regime we will be able to use the Ionescu–Wainger multiplier theorem (see Lemma 5.2(iv) and Remark 5.11 below) as a partial replacement[7] of this transference, at least

[7] Another partial replacement of Theorem 4.18 in this setting was recently established in [80, Th. 1.6].

at the level of linear Fourier multiplier operators. The Ionescu–Wainger theory does not directly treat the "twisted" bilinear multipliers $\mathrm{B}_m^{l_1,l_2,m_{\hat{\mathbb{Z}}}}$ that we will eventually need to handle (see (7.10)), so we will need to first apply a two-parameter Rademacher–Menshov argument in order to reduce the bilinear analysis to linear estimates that can be treated by that theory; see Section 8.

We close this section with some crude multiplier estimates on $\mathbb{Z}$ and on $\mathbb{R}$.

LEMMA 4.20 (Crude multiplier bound). *Let $\mathbb{G} = \mathbb{Z}$ or $\mathbb{G} = \mathbb{R}$.*

(i) *Let $\varphi \in \boldsymbol{S}(\mathbb{G}^*)$ and $r > 0$. When $\mathbb{G} = \mathbb{Z}$, we also require $r \leq 1$. Then for any $1 \leq p \leq \infty$, T_φ extends continuously to a linear map from $L^p(\mathbb{G})$ to $L^p(\mathbb{G})$ with*

$$\|\mathrm{T}_\varphi\|_{L^p(\mathbb{G})\to L^p(\mathbb{G})} \lesssim \sup_{0\leq j\leq 2} \int_{\mathbb{G}^*} r^{j-1} \left|\frac{d^j}{d\xi^j}\varphi(\xi)\right| \, d\xi.$$

(ii) *Let $m \in \boldsymbol{S}(\mathbb{G}^* \times \mathbb{G}^*)$, $r_1, r_2 > 0$, and $1 \leq p, p_1, p_2 \leq \infty$ with $\frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{p}$. When $\mathbb{G} = \mathbb{Z}$ we also require $r_1, r_2 \leq 1$. Then B_m extends continuously to a bilinear map from $L^{p_1}(\mathbb{G}) \times L^{p_2}(\mathbb{G})$ to $L^p(\mathbb{G})$ with*

$$\|\mathrm{B}_m\|_{L^{p_1}(\mathbb{G})\times L^{p_2}(\mathbb{G})\to L^p(\mathbb{G})}$$

$$\lesssim \sup_{0\leq j_1,j_2\leq 2} \int_{\mathbb{G}^*}\int_{\mathbb{G}^*} r_1^{j_1-1} r_2^{j_2-1} \left|\frac{\partial^{j_1}}{\partial\xi_1^{j_1}}\frac{\partial^{j_2}}{\partial\xi_1^{j_2}} m(\xi_1,\xi_2)\right| \, d\xi_1 d\xi_2. \tag{4.21}$$

The same bound also holds when the hypothesis $1 \leq p, p_1, p_2 \leq \infty$ is replaced by $1 < p_1, p_2 \leq \infty$, except now the implied constant in (4.21) is permitted to depend on p_1, p_2.

Proof. We just prove (ii) in the case $\mathbb{G} = \mathbb{Z}$, as all the other cases are similar. It suffices to prove the claim for Schwartz functions. We may normalize the right-hand side of (4.21) to be 1. We can express B_m in physical space as

$$\mathrm{B}_m(f,g)(x) = \sum_{y_1,y_2\in\mathbb{Z}} K(y_1,y_2) f(x-y_1) g(x-y_2),$$

where

$$K(y_1,y_2) := \int_{\mathbb{T}^2} m(\xi_1,\xi_2) e(-y_1\xi_1 - y_2\xi_2) \, d\xi_1\xi_2.$$

Suppose first that we are in the case $1 \leq p, p_1, p_2 \leq \infty$. By Minkowski's inequality, we have

$$\|\mathrm{B}_m\|_{\ell^{p_1}(\mathbb{Z})\times\ell^{p_2}(\mathbb{Z})\to\ell^p(\mathbb{Z})} \leq \|K\|_{\ell^1(\mathbb{Z}^2)}.$$

On the other hand, from the normalization of (4.21) and integration by parts, we have

$$K(y_1,y_2) \lesssim r_1^{1-j_1} r_2^{1-j_2} |y_1|^{-j_1} |y_2|^{-j_2}$$

for any $y_1, y_2 \in \mathbb{Z}$ and $0 \leq j_1, j_2 \leq 2$ (with the claim being vacuously true if the right-hand side is infinite), thus

$$K(y_1,y_2) \lesssim r_1\langle r_1 y_1\rangle^{-2} r_2 \langle r_2 y_2\rangle^{-2}, \tag{4.22}$$

and the claim follows. In the case $1 < p_1, p_2 \leq \infty$, we can instead use (4.22) to bound $\mathrm{B}_m(f,g)$ pointwise by the product of the Hardy–Littlewood maximal functions of f, g, and the claim now follows from Hölder's inequality and the Hardy–Littlewood maximal inequality. □

5. **Ionescu–Wainger decomposition: reducing to major arcs**

We now begin the proof of Theorem 3.9. Henceforth the parameters $P, d, p_1, p_2, p, r, \lambda$ are fixed to obey the hypotheses of this theorem, and all implied constants in the asymptotic notation are allowed to depend on these parameters. We also fix the finite λ-lacunary subset $\mathbb{D}$ of $\mathbb{Z}_+$, although we require all our estimates to be uniform in the choice of $\mathbb{D}$. We abbreviate $\tilde{A}_N^{\mathrm{n},P(\mathrm{n})}$ as $\tilde{A}_N$.

We will also need four large constants:

(i) We choose a constant $C_0 \in \mathbb{Z}_+$ that is sufficiently large depending on the fixed parameters $P, d, p_1, p_2, p, r, \lambda$. (This constant is used to define a maximum height scale $l_{(N)}$ associated to each physical scale N; see (5.22).)

(ii) We choose a constant $C_1 \in \mathbb{Z}_+$ that is sufficiently large depending on the fixed parameters $P, d, p_1, p_2, p, r, \lambda$ and on C_0. (This constant is used to define the Ionescu–Wainger parameter ρ; see (5.1).)

(iii) We choose a constant $C_2 \in \mathbb{Z}_+$ that is sufficiently large depending on the fixed parameters $P, d, p_1, p_2, p, r, \lambda$ and on C_0, C_1. (This quantity is used to define an auxiliary scale u associated to a given height scale l; see (5.26).)

(iv) We choose a constant $C_3 \in \mathbb{Z}_+$ that is sufficiently large depending on the fixed parameters $P, d, p_1, p_2, p, r, \lambda$ and on C_0, C_1, C_2. (This quantity will be used to lower bound the physical scale N, as well as to bound implied constants in estimates.)

We also use $c > 0$ to denote various small exponents that depend only on d, p_1, p_2, p, r, and which will vary from line to line. Occasionally we will also need c to depend on some other parameters, and we will indicate this by additional subscripts; for instance, c_q will be a positive constant depending on d, p_1, p_2, p, r, q. Importantly, these constants c will *not* depend on the large constants C_0, C_1, C_2, C_3 just introduced. Specifically, c will be independent on the Ionescu–Wainger parameter ρ; see (5.1).

Define the *naive height* $\mathrm{h}_{\mathrm{naive}}(\alpha) \in 2^{\mathbb{N}}$ of an arithmetic frequency $\alpha = \frac{a}{q} \bmod 1 \in \mathbb{Q}/\mathbb{Z}$ by the formula

$$\mathrm{h}_{\mathrm{naive}}\left(\frac{a}{q} \bmod 1\right) := \inf\{2^l : l \in \mathbb{N}, q \leq 2^l\} = 2^{\lceil \log q \rceil} \sim q$$

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

$$(\mathbb{Q}/\mathbb{Z})_{\leq l,\mathrm{naive}} := \mathrm{h}_{\mathrm{naive}}^{-1}([2^l]) = \{\alpha \in \mathbb{Q}/\mathbb{Z}\colon\ \mathrm{h}_{\mathrm{naive}}(\alpha) \leq 2^l\}$$

and the continuous frequency sets

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

and then define the naive major arcs

$$\mathcal{M}_{\leq l,\leq k,\mathrm{naive}} := \pi(\mathbb{R}_{\leq k} \times (\mathbb{Q}/\mathbb{Z})_{\leq l,\mathrm{naive}}).$$

Thus $\mathcal{M}_{\leq l,\leq k,\mathrm{naive}}$ consists of all elements of $\mathbb{T}$ of the form $\frac{a}{q} + \theta \bmod 1$ for some $q \in [2^l]$, $a \in [q]^\times$, and $\theta \in [-2^{-k}, 2^k]$. These would be the obvious choice of major arcs to restrict attention to in our Fourier-analytic manipulations. Unfortunately, the L^p multiplier theory on such arcs is unfavorable. To obtain a better theory, we follow Ionescu and Wainger [47] and replace the naive height $\mathrm{h}_{\mathrm{naive}}(\alpha)$ of an arithmetic frequency by a smaller quantity, which we call the *Ionescu–Wainger height* $\mathrm{h}(\alpha) = \mathrm{h}_\rho(\alpha) \in 2^{\mathbb{N}}$. This height depends on an additional small parameter $0 < \rho < 1$, which we now fix in our hierarchy of constants as

$$\rho := 1/C_1. \tag{5.1}$$

The precise definition of this height is technical and is postponed to Appendix A. However, for our purposes we can summarize the main properties of this height as follows. Using this height, we define the Ionescu–Wainger arithmetic frequency sets

$$(\mathbb{Q}/\mathbb{Z})_{\leq l} := \mathrm{h}^{-1}([2^l]) = \{\alpha \in \mathbb{Q}/\mathbb{Z}\colon\ \mathrm{h}(\alpha) \leq 2^l\}$$

and the Ionescu–Wainger major arcs or simply major arcs

$$\mathcal{M}_{\leq l,\leq k} := \pi(\mathbb{R}_{\leq k} \times (\mathbb{Q}/\mathbb{Z})_{\leq l});$$

see Figure 6. These arcs will be somewhat larger than their naive counterparts, but this is more than compensated for by their superior Fourier multiplier theory. We also use the variants

$$(\mathbb{Q}/\mathbb{Z})_l := (\mathbb{Q}/\mathbb{Z})_{\leq l} \backslash (\mathbb{Q}/\mathbb{Z})_{\leq l-1} = \mathrm{h}^{-1}(2^l) = \{\alpha \in \mathbb{Q}/\mathbb{Z}\colon\ \mathrm{h}(\alpha) = 2^l\}$$

and

$$\mathcal{M}_{l,\leq k} := \pi(\mathbb{R}_{\leq k} \times (\mathbb{Q}/\mathbb{Z})_l)$$

with the convention that $(\mathbb{Q}/\mathbb{Z})_{\leq -1}$ is empty.

LEMMA 5.2 (Properties of height).

(i) (Naive height controls height). *For any $\alpha \in \mathbb{Q}/\mathbb{Z}$, one has*

$$\mathrm{h}(\alpha) \leq \mathrm{h}_{\mathrm{naive}}(\alpha). \tag{5.3}$$

In particular, $(\mathbb{Q}/\mathbb{Z})_{\leq l,\mathrm{naive}} \subset (\mathbb{Q}/\mathbb{Z})_{\leq l}$ and $\mathcal{M}_{\leq l,\leq k,\mathrm{naive}} \subset \mathcal{M}_{\leq l,\leq k}$ for any $(l,k) \in \mathbb{N}\times\mathbb{Z}$. If $\alpha \in \frac{1}{p}\mathbb{Z}/\mathbb{Z}$ for a prime p, then equality holds in (5.3).

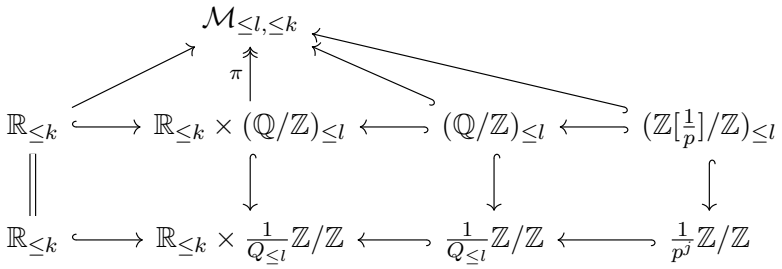


Figure 6. *The commutative diagram in Figure 2, restricted to major arcs, where p^j is the largest power of p dividing $Q_{\leq l}$ and $(\mathbb{Z}[\frac{1}{p}]/\mathbb{Z})_{\leq l} := (\mathbb{Q}/\mathbb{Z})_{\leq l} \cap \mathbb{Z}[\frac{1}{p}]/\mathbb{Z}$.* When (l,k) has good major arcs, the product set $\mathbb{R}_{\leq k} \times (\mathbb{Q}/\mathbb{Z})_{\leq l}$ is non-aliasing, and the indicated map π can be upgraded from a surjection to a bijection. Most of the spaces in this diagram are no longer groups and so the arrows are now downgraded from continuous homomorphisms to continuous maps. Note the approximate duality with Figure 3.

(ii) (Cyclic structure). *For any $l \in \mathbb{N}$, $(\mathbb{Q}/\mathbb{Z})_{\leq l}$ is the union of finitely many dual cyclic groups $\frac{1}{q}\mathbb{Z}/\mathbb{Z}$ with*

$$q \lesssim_\rho 2^{2^{\rho l}}$$

and is contained in a single dual cyclic group $\frac{1}{Q_{\leq l}}\mathbb{Z}/\mathbb{Z}$ with

$$Q_{\leq l} \lesssim 2^{O_\rho(2^l)}.$$

In fact, the integer $Q_{\leq l} \in \mathbb{Z}_+$ can be defined explicitly as in (A.4).

(iii) (Cardinality bound). *For any $l \in \mathbb{N}$, one has*

$$\#(\mathbb{Q}/\mathbb{Z})_{\leq l} \lesssim_\rho 2^{2^{\rho l}}.$$

Proof. See Appendix A. □

The linear Fourier multiplier operators $\mathrm{T}_\varphi^{\leq l}$ and T_φ^l defined by

$$\mathrm{T}_\varphi^{\leq l} := \mathrm{T}_{\mathcal{P}(\varphi \otimes \mathbb{1}_{(\mathbb{Q}/\mathbb{Z})_{\leq l}})}, \tag{5.4}$$

$$\mathrm{T}_\varphi^{l} := \mathrm{T}_{\mathcal{P}(\varphi \otimes \mathbb{1}_{(\mathbb{Q}/\mathbb{Z})_{l}})}, \tag{5.5}$$

will play a key role in our analysis. They can be written more explicitly as

$$\mathrm{T}_\varphi^{\leq l} f(x) = \sum_{\alpha \in (\mathbb{Q}/\mathbb{Z})_{\leq l}} \int_{\mathbb{R}} \varphi(\theta) \mathcal{F}_{\mathbb{Z}} f(\alpha+\theta) e(-x(\alpha+\theta))\, d\theta,$$

$$\mathrm{T}_\varphi^{l} f(x) = \sum_{\alpha \in (\mathbb{Q}/\mathbb{Z})_{l}} \int_{\mathbb{R}} \varphi(\theta) \mathcal{F}_{\mathbb{Z}} f(\alpha+\theta) e(-x(\alpha+\theta))\, d\theta.$$

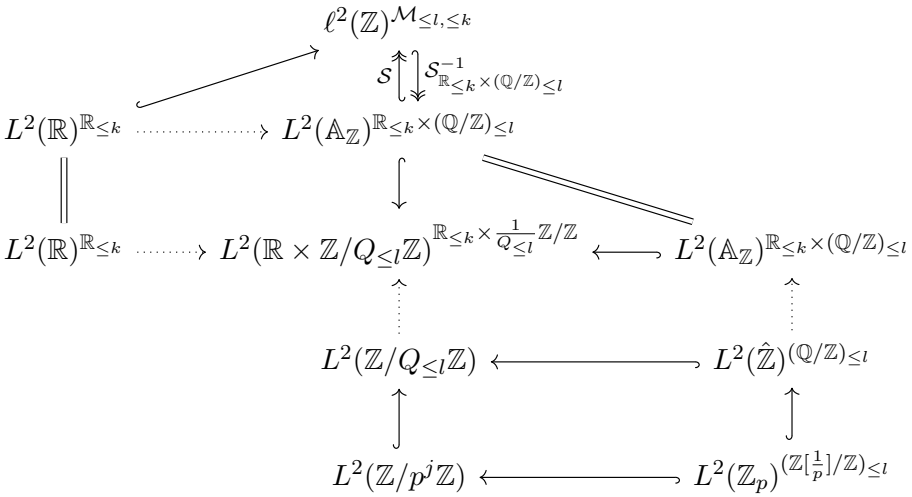


Figure 7. The L^2 version of Figure 6, under the hypothesis of good major arcs. Solid (hooked) arrows are Hilbert space isometries, double-headed arrows are unitary maps, and dotted arrows indicate a (Hilbert space) tensor product. We thus see that the major arc component $\ell^2(\mathbb{Z})^{\mathcal{M}_{\leq l,\leq k}}$ of $\ell^2(\mathbb{Z})$ can be identified with the tensor product of the low (continuous) frequency component $L^2(\mathbb{R})^{\mathbb{R}_{\leq k}}$ of $L^2(\mathbb{R})$ and the low (arithmetic) frequency component $L^2(\hat{\mathbb{Z}})^{(\mathbb{Q}/\mathbb{Z})_{\leq l}}$ of $L^2(\hat{\mathbb{Z}})$, with the latter component identifiable in turn with a subspace of $L^2(\mathbb{Z}/Q_{\leq l}\mathbb{Z})$. As with Figure 4, some arrows are missing due to the failure of $L^2(\mathbb{R})^{\mathbb{R}_{\leq k}}$ to contain a unit 1.

From (4.5) and (4.9), one has the functional calculus

$$\mathrm{T}^{\leq l}_{\varphi_1\varphi_2} = \mathrm{T}^{\leq l}_{\varphi_1} \circ T^{\leq l}_{\varphi_2} \tag{5.6}$$

whenever (l,k) has good major arcs and $\varphi_1, \varphi_2 \in \boldsymbol{S}(\mathbb{R}_{\leq k})$ — and similarly with $\leq l$ replaced by l in (5.6). The principal tool in bounding operators (5.4) and (5.5) is the Ionescu–Wainger multiplier theorem [47], which for our purposes can be formulated as follows:

THEOREM 5.7 (Vector-valued Ionescu–Wainger multiplier theorem). *If* $(l,k) \in \mathbb{N}\times\mathbb{Z}$ *has* good major arcs *in the sense that*

$$k \leq -C_\rho 2^{\rho l} \tag{5.8}$$

for a sufficiently large constant C_ρ *depending only on* ρ*, then the compact set* $\mathbb{R}_{\leq k}\times(\mathbb{Q}/\mathbb{Z})_{\leq l} \subset \mathbb{R}\times\mathbb{Q}/\mathbb{Z}$ *is non-aliasing. Furthermore, if* $q \in 2\mathbb{N}\cup(2\mathbb{N})'$ *is either an even integer or the dual of an even integer, then the linear Fourier*

multiplier operator $\mathrm{T}^{\leq l}_{\vec{\varphi}}$ *from* (5.4) *obeys the multiplier bound*

$$\|\mathrm{T}^{\leq l}_{\vec{\varphi}}\|_{\ell^q(\mathbb{Z};H)\to\ell^q(\mathbb{Z};H)} \lesssim_{\rho,q} \langle l\rangle \|\mathrm{T}_{\varphi}\|_{L^q(\mathbb{R})\to L^q(\mathbb{R})} \tag{5.9}$$

for any $\varphi \in \boldsymbol{S}(\mathbb{R}_{\leq k})$, *and any finite-dimensional Hilbert space* H, *and similarly for the multiplier operator* T^{l}_{φ} *from* (5.5).

Proof. See Appendix A. □

Remark 5.10. Some remarks about Theorem 5.7 are in order.

(i) Theorem 5.7 in the scalar-valued setting was first established by Ionescu and Wainger [47] with the factor $\langle l\rangle^{\lfloor 2/\rho\rfloor+1}$ in place of $\langle l\rangle$ in (5.9). Their proof is based on an intricate inductive argument that exploits super-orthogonality phenomena. A slightly different proof (giving the factor $\langle l\rangle$ in (5.9)) using certain recursive arguments, which clarified the role of the underlying square functions and orthogonalities, was presented in [65]. A vector-valued Ionescu–Wainger multiplier theorem (in the spirit of [65]) can be found in [68, §2]. A uniform vector-valued Ionescu–Wainger multiplier theorem, where the factor $\langle l\rangle$ is removed from (5.9), was recently proved by the third author [80]. The latter proof also provides explicit constants in (5.9) and allows us to handle adelic Fourier multipliers. The super-orthogonality phenomena are discussed in the survey of Pierce [76] in a much broader context.

(ii) The fact that the losses in (5.9) are only polynomial in the logarithmic height scale l instead of exponential will be essential to our arguments, and form the main reason why we cannot work with the naive notion of heights, as the analogous multiplier theorem is not available for such heights.

(iii) As we are focused on variational estimates, even the factors like $2^{O(\rho l)}$ will have to be handled; see the constants produced by the Rademacher–Menshov inequality in Section 8. From this point of view, even though the uniform vector-valued Ionescu–Wainger multiplier theorem [80] is now available, and the factor $\langle l\rangle$ can be deleted, this does not significantly improve the main result or simplify the proof. Hence, we will use the vector-valued Ionescu–Wainger multiplier theorem from [68, §2].

(iv) The restriction in Theorem 5.7 to the case when q is an even integer or the dual of an even integer can be ignored in practice because in all the applications of Theorem 5.7 we will have good $L^q(\mathbb{R})$ operator norm bounds on T_{φ} for *all* $1 < q < \infty$. Then by applying (5.9) for $q \in 2\mathbb{N}\cup(2\mathbb{N})'$ and then interpolating, we can recover good bounds for all $1 < q < \infty$. See also the discussion after [68, Th. 2.1].

Remark 5.11. When (l,k) has good major arcs, the corresponding sampling operator $\mathcal{S}\colon L^2(\mathbb{A}_{\mathbb{Z}})^{\mathbb{R}_{\leq k}\times(\mathbb{Q}/\mathbb{Z})_{\leq l}} \to \ell^2(\mathbb{Z})^{\mathcal{M}_{\leq l,\leq k}}$ is unitary thanks to (4.6), and it is inverted by the interpolation operator $\mathcal{S}^{-1}_{\mathbb{R}_{\leq k}\times(\mathbb{Q}/\mathbb{Z})_{\leq l}}$; see Figure 7. For L^p norms, one no longer expects to have the isometry property even at an approximate level (except in the large scale case when Theorem 4.18 applies), but (5.9) shows that at least the linear Fourier multiplier theory on $\ell^q(\mathbb{Z})^{\mathcal{M}_{\leq l,\leq k}}$ is basically controlled (up to small losses) by that of $L^q(\mathbb{A}_{\mathbb{Z}})^{\mathbb{R}_{\leq k}\times(\mathbb{Q}/\mathbb{Z})_{\leq l}}$ (at least when $q \in (2\mathbb{N}) \cup (2\mathbb{N})'$), which serves as a partial substitute for an isometry property for the sampling operator.

A crucial component of our arguments is the assertion that the bilinear averaging operator $\tilde{A}_N^{\mathrm{n},P(\mathrm{n})}(f,g)$ is negligible when the Fourier transform of f or g vanishes on major arcs. More precisely, we have the following improvement of (1.6) in this case:

THEOREM 5.12 (Single scale minor arc estimate). *Let $N \geq 1$, let $l \in \mathbb{N}$, and suppose that $f, g \in \ell^2(\mathbb{Z})$ obeys one of the following assumptions*:

(i) *$\mathcal{F}_{\mathbb{Z}} f$ vanishes on $\mathcal{M}_{\leq l,\leq -\operatorname{Log} N + l}$;*
(ii) *$\mathcal{F}_{\mathbb{Z}} g$ vanishes on $\mathcal{M}_{\leq l,\leq -d\operatorname{Log} N + dl}$,*

where the logarithmic scale $\operatorname{Log} N$ of N was defined in (2.1). Then one has

$$\|\tilde{A}_N(f,g)\|_{\ell^1(\mathbb{Z})} \lesssim_{C_1} (2^{-cl} + \langle \operatorname{Log} N\rangle^{-cC_1})\|f\|_{\ell^2(\mathbb{Z})}\|g\|_{\ell^2(\mathbb{Z})}. \tag{5.13}$$

This theorem will be used repeatedly in our arguments. The parameter $c > 0$ from (5.13) will be independent on the Ionescu–Wainger parameter ρ; see (5.1). The secondary term $\langle \operatorname{Log} N\rangle^{-cC_1}$ is negligible in practice; the key point is the primary term 2^{-cl} that exhibits exponential decay on the height scale l. It is important to note that only one of the hypotheses (i), (ii), as opposed to both, are required to hold in order to obtain this decay. The asymmetry between (i) and (ii) is entirely caused by the different degrees in the two polynomials $\mathrm{n}, P(\mathrm{n})$ used to form the averaging operator $\tilde{A}_N$. This theorem only gives exponential decay directly for $\ell^2(\mathbb{Z})\times\ell^2(\mathbb{Z}) \to \ell^1(\mathbb{Z})$ operator norms, but in practice one can use interpolation to then obtain similar decay for other $\ell^{p_1}(\mathbb{Z}) \times \ell^{p_2}(\mathbb{Z}) \to \ell^p(\mathbb{Z})$ operator norms. We remark that it is essential in Theorem 5.12 that we are in the nonlinear regime $d \geq 2$, as there are easy counterexamples to this theorem in the linear case $d = 1$ (as can be seen by testing (5.13) against plane waves multiplied by suitable cutoff functions).

The proof of Theorem 5.12 will be somewhat lengthy, and it relies on several deep results in the literature, including the inverse theory of Peluse and Prendiville [74] and Peluse [73] (see also [75] and the survey of Prendiville [77]) and L^p-improving estimates of Han–Kovač–Lacey–Madrid–Yang [41] (see also Dasu–Demeter–Langowski [20]); we also use the properties of the Ionescu–Wainger projections that we shall define later in this section. A key difficulty in

the proof of Theorem 5.12 will be that the functions f, g are only controlled in $\ell^2(\mathbb{Z})$ rather than $\ell^\infty(\mathbb{Z})$. We will establish this bound in Section 6. We remark that a continuous analogue of Theorem 5.12, with the domain $\mathbb{Z}$ replaced by $\mathbb{R}$, and with the major arc set replaced by an interval centered at the frequency origin, was established in [10, Lemma 5] for monomial P and in [28, Lemma 1.4] in the general case.

Example 5.14. Let $l \in \mathbb{N}$, and let N be a sufficiently large integer depending on l, P. Let q be a prime number with $2^l < q \leq 2^{l+1}$ (which implies, in particular, that $\frac{1}{q} \bmod 1$ has height 2^{l+1}). Consider the functions

$$f(n) \coloneqq e(-n/q) \sum_{j \in [N^{d-1}]} \epsilon_j (\mathcal{F}_{\mathbb{R}}^{-1}\eta)\left(\frac{n - jN}{N}\right),$$

$$g(n) \coloneqq e(-n/q)(\mathcal{F}_{\mathbb{R}}^{-1}\eta)\left(\frac{n}{N^d}\right),$$

where $\epsilon_1, \dots, \epsilon_{N^{d-1}} \in \{-1, +1\}$ are arbitrary signs and η is defined in Section 2.3. Then $\mathcal{F}_{\mathbb{Z}} f$ and $\mathcal{F}_{\mathbb{Z}} g$ vanish on $\mathcal{M}_{\leq l, \leq -\operatorname{Log} N + l}$ and $\mathcal{M}_{\leq l, \leq -d \operatorname{Log} N + dl}$ respectively, and routine calculations show that

$$\|\mathcal{F}_{\mathbb{Z}} f\|_{\ell^2(\mathbb{Z})}, \|\mathcal{F}_{\mathbb{Z}} g\|_{\ell^2(\mathbb{Z})} \lesssim N^{d/2}$$

and also

$$\|\tilde{A}_N(f, g)\|_{\ell^1(\mathbb{Z})} \lesssim N^d \left(\left| \mathbb{E}_{n \in \mathbb{Z}/q\mathbb{Z}} e\left(\frac{n + P(n)}{q}\right) \right| + N^{-c} \right).$$

Standard exponential sum estimates (see, e.g., [48]) reveal that

$$\mathbb{E}_{n \in \mathbb{Z}/q\mathbb{Z}} e\left(\frac{n + P(n)}{q}\right) \lesssim q^{-c} \lesssim 2^{-cl}$$

(indeed, the Weil bounds allow one to take $c = 1/2$ here), and so this example is consistent with Theorem 5.12. Variations of this example can also be used to explain the appearance of the scales $-\operatorname{Log} N$ and $-d \operatorname{Log} N$ in Theorem 5.12(i), (ii), which are the frequency dual scales to the spatial scales $\operatorname{Log} N$, $\operatorname{Log} N^d$ associated to the shifts $n, P(n)$ for $n \in [N]$ arising in the definition of $A_N^{\mathrm{n}, P(\mathrm{n})}$; we leave the details to the interested reader.

For the remainder of this section, let us assume Theorem 5.12 and see how we can use it to attack Theorem 3.9. We will need an adelic version of Littlewood–Paley projection operators. Let $\eta_{\leq k}$ be the cutoff functions from Section 2.3. The Fourier multipliers $\mathrm{T}_{\eta_{\leq k}}$ are then standard Littlewood–Paley Fourier projections on $\boldsymbol{S}(\mathbb{R})$ to the frequency interval $\mathbb{R}_{\leq k}$. Motivated by this, we define the *Ionescu–Wainger Fourier projection operator* $\Pi_{\leq l, \leq k}$ for any $(l, k) \in \mathbb{N} \times \mathbb{Z}$ using the construction (5.4) by the formula

$$\Pi_{\leq l, \leq k} \coloneqq \mathrm{T}^{\leq l}_{\eta_{\leq k}}. \tag{5.15}$$

More explicitly, one has

$$\Pi_{\leq l,\leq k} f(x) = \sum_{\alpha \in (\mathbb{Q}/\mathbb{Z})_{\leq l}} \int_{\mathbb{R}} \eta(\theta/2^k) \mathcal{F}_{\mathbb{Z}} f(\alpha+\theta) e(-x(\alpha+\theta))\, d\theta.$$

Note that $\Pi_{\leq l,\leq k}$ is self-adjoint on $\ell^2(\mathbb{Z})$, and its symbol is supported on $\mathcal{M}_{\leq l,\leq k}$. We similarly define

$$\Pi_{l,\leq k} := \mathrm{T}^l_{\eta_{\leq k}} = \Pi_{\leq l,\leq k} - \Pi_{\leq l-1,\leq k} \tag{5.16}$$

with the convention $\Pi_{\leq -1,k} = 0$.

When (l,k) have good major arcs, these operators have good properties:

LEMMA 5.17 (Properties of Ionescu–Wainger projections). *Let $(l,k) \in \mathbb{N}\times\mathbb{Z}$ be such that (l,k) has good major arcs.*

(i) (Boundedness). *The operator $\Pi_{\leq l,\leq k}$ is a contraction on $\ell^2(\mathbb{Z})$. Furthermore, for any $1<q<\infty$, one has*

$$\|\Pi_{\leq l,\leq k} f\|_{\ell^q(\mathbb{Z})} \lesssim_{C_1,q} \langle l\rangle \|f\|_{\ell^q(\mathbb{Z})}. \tag{5.18}$$

In particular, $\Pi_{\leq l,\leq k}$ extends to a bounded linear operator on $\ell^q(\mathbb{Z})$. If f is furthermore supported on an interval I, we have the off-diagonal decay bound

$$\|\Pi_{\leq l,\leq k} f\|_{\ell^q(J)} \lesssim_{C_1,q,M} \langle l\rangle \langle 2^k \mathrm{dist}(I,J)\rangle^{-M} \|f\|_{\ell^q(I)} \tag{5.19}$$

for any interval J, and any $M\in\mathbb{N}$.

(ii) (Fourier support). *If $f\in\ell^2(\mathbb{Z})$, then $\Pi_{\leq l,\leq k} f$ is Fourier supported in $\mathcal{M}_{\leq l,\leq k}$, and $\Pi_{\leq l,\leq k} f = f$ when $\mathcal{F}_{\mathbb{Z}} f$ is Fourier supported in $\mathcal{M}_{\leq l,\leq k-1}$.*

All these claims also hold when all occurrences of $\leq l$ are replaced by l.

Proof. See Appendix A. □

Remark 5.20 (Physical space interpretation of major arcs). By uncertainty principle heuristics, functions $f\in \boldsymbol{S}(\mathbb{Z})$ that have Fourier support in $\mathcal{M}_{\leq l,\leq k}$, where $(l,k)\in\mathbb{N}\times\mathbb{Z}$ satisfy (5.8), can be viewed as behaving like linear combinations of indicator functions $\mathbb{1}_P$ of arithmetic progressions P of spacing $O(2^l)$ and diameter $O(2^{-k})$, and they behave like constants on arithmetic progressions of spacing $Q_{\leq l}$ and diameter $O(2^{-k})$; the latter is only non-vacuous in the "large-scale" regime in which 2^{-k} is larger than $Q_{\leq l}$. Dually, functions $f\in\boldsymbol{S}(\mathbb{Z})$ whose Fourier transform vanishes on $\mathcal{M}_{\leq l,\leq k}$ morally have negligible mean on the two types of arithmetic progressions just mentioned. The reader is invited to compare Figure 3 with Figure 6 through the lens of this uncertainty principle.

Now we can use Theorem 5.12 and Lemma 5.17 to achieve some reductions to prove Theorem 3.9. It will suffice to establish the estimate

$$\|(\tilde{A}_N(f,g))_{N\in\mathbb{D}}\|_{\ell^p(\mathbb{Z};\boldsymbol{V}^r)} \lesssim_{C_3} \|f\|_{\ell^{p_1}(\mathbb{Z})}\|g\|_{\ell^{p_2}(\mathbb{Z})}. \tag{5.21}$$

For each individual $N < C_3$, this claim is immediate from (1.6), so we may assume without loss of generality that $N \geq C_3$ for all $N \in \mathbb{D}$. If $N \geq C_3$, define the quantities

$$l_{(N)} := C_0 \operatorname{Log}\operatorname{Log} N. \tag{5.22}$$

Then by (5.8), the pairs $(l_{(N)}, -\operatorname{Log} N + l_{(N)})$, $(l_{(N)}, -d\operatorname{Log} N + dl_{(N)})$ have good major arcs, and hence by Lemma 5.17(i), (ii) and Theorem 5.12, if $C_1 \geq C_0$, one has the estimate

$$\|\tilde{A}_N((1-\Pi_{\leq l_{(N)},-\operatorname{Log} N+l_{(N)}})f,g)\|_{\ell^1(\mathbb{Z})} \lesssim_{C_1} (\operatorname{Log} N)^{-cC_0}\|f\|_{\ell^2(\mathbb{Z})}\|g\|_{\ell^2(\mathbb{Z})}.$$

On the other hand, from Lemma 5.17(i) and (1.6) one also has

$$\|\tilde{A}_N((1-\Pi_{\leq l_{(N)},-\operatorname{Log} N+l_{(N)}})f,g)\|_{\ell^q(\mathbb{Z})} \lesssim_{C_1,q,q_1,q_2} (\operatorname{Log}\operatorname{Log} N)\|f\|_{\ell^{q_1}(\mathbb{Z})}\|g\|_{\ell^{q_2}(\mathbb{Z})}$$

for any $1 < q_1, q_2 < \infty$ with $1/q_1 + 1/q_2 = 1/q \leq 1$. Interpolating, we conclude that

$$\|\tilde{A}_N((1-\Pi_{\leq l_{(N)},-\operatorname{Log} N+l_{(N)}})f,g)\|_{\ell^p(\mathbb{Z})} \lesssim_{C_1} (\operatorname{Log}\operatorname{Log} N)^{O(1)}(\operatorname{Log} N)^{-cC_0}\|f\|_{\ell^{p_1}(\mathbb{Z})}\|g\|_{\ell^{p_2}(\mathbb{Z})}.$$

(Recall that c varies from line to line and is allowed to depend on p_1, p_2, p.) In particular, for C_0 large enough, one has (say)

$$\|\tilde{A}_N((1-\Pi_{\leq l_{(N)},-\operatorname{Log} N+l_{(N)}})f,g)\|_{\ell^p(\mathbb{Z})} \lesssim_{C_1} (\operatorname{Log} N)^{-10}\|f\|_{\ell^{p_1}(\mathbb{Z})}\|g\|_{\ell^{p_2}(\mathbb{Z})}.$$

A similar argument gives

$$\|\tilde{A}_N(\Pi_{\leq l_{(N)},-\operatorname{Log} N+l_{(N)}}f,(1-\Pi_{\leq l_{(N)},\leq -d\operatorname{Log} N+dl_{(N)}})g)\|_{\ell^p(\mathbb{Z})} \lesssim_{C_1} (\operatorname{Log} N)^{-10}\|f\|_{\ell^{p_1}(\mathbb{Z})}\|g\|_{\ell^{p_2}(\mathbb{Z})};$$

by the triangle inequality and bilinearity of $\tilde{A}_N$, we conclude that

$$\|\tilde{A}_N(f,g) - \tilde{A}_N(\Pi_{\leq l_{(N)},\leq -\operatorname{Log} N+l_{(N)}}f,\Pi_{\leq l_{(N)},\leq -d\operatorname{Log} N+dl_{(N)}}g)\|_{\ell^p(\mathbb{Z})} \lesssim_{C_1} (\operatorname{Log} N)^{-10}\|f\|_{\ell^{p_1}(\mathbb{Z})}\|g\|_{\ell^{p_2}(\mathbb{Z})}.$$

From the λ-lacunary nature of $\mathbb{D}$, we have

$$\sum_{N\in\mathbb{D}:N\geq C_3} (\operatorname{Log} N)^{-10} \lesssim 1$$

and hence by (2.7), we have that

$$\|(\tilde{A}_N(f,g) - \tilde{A}_N(\Pi_{\leq l_{(N)},\leq -\operatorname{Log} N + l_{(N)}} f, \Pi_{\leq l_{(N)},\leq -d\operatorname{Log} N + dl_{(N)}} g))_{N\in\mathbb{D}}\|_{\ell^p(\mathbb{Z};\boldsymbol{V}^r)} \lesssim_{C_1} \|f\|_{\ell^{p_1}(\mathbb{Z})}\|g\|_{\ell^{p_2}(\mathbb{Z})}.$$

By a further application of the triangle inequality, we conclude that to establish (5.21), it suffices to prove the major arc bound

$$\|(\tilde{A}_N(\Pi_{\leq l_{(N)},\leq -\operatorname{Log} N + l_{(N)}} f, \Pi_{\leq l_{(N)},\leq -d\operatorname{Log} N + dl_{(N)}} g))_{N\in\mathbb{D}}\|_{\ell^p(\mathbb{Z};\boldsymbol{V}^r)} \lesssim_{C_3} \|f\|_{\ell^{p_1}(\mathbb{Z})}\|g\|_{\ell^{p_2}(\mathbb{Z})}.$$

We now perform an "arithmetic" dyadic decomposition

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

By the triangle inequality, it now suffices to show the bound

$$\|(\tilde{A}_N(\Pi_{l_1,\leq -\operatorname{Log} N + l_{(N)}} f, \Pi_{l_2,\leq -d\operatorname{Log} N + dl_{(N)}} g)\mathbb{1}_{l_1,l_2\leq l_{(N)}})_{N\in\mathbb{D}}\|_{\ell^p(\mathbb{Z};\boldsymbol{V}^r)} \lesssim_{C_3} 2^{-\rho l}\|f\|_{\ell^{p_1}(\mathbb{Z})}\|g\|_{\ell^{p_2}(\mathbb{Z})} \tag{5.23}$$

for all $l_1, l_2 \in \mathbb{N}$, where

$$l := \max(l_1, l_2). \tag{5.24}$$

Note that the constraint $l_1, l_2 \leq l_{(N)}$ serves as an additional lower bound on N (and, in particular, the left-hand side of (5.23) vanishes for all but finitely many l_1, l_2, thanks to the finite nature of $\mathbb{D}$), so we may also write this bound as

$$\|(\tilde{A}_N(\Pi_{l_1,\leq -\operatorname{Log} N + l_{(N)}} f, \Pi_{l_2,\leq -d\operatorname{Log} N + dl_{(N)}} g))_{N\in\mathbb{D};l_1,l_2\leq l_{(N)}}\|_{\ell^p(\mathbb{Z};\boldsymbol{V}^r)} \lesssim_{C_3} 2^{-\rho l}\|f\|_{\ell^{p_1}(\mathbb{Z})}\|g\|_{\ell^{p_2}(\mathbb{Z})}. \tag{5.25}$$

Fix l_1, l_2 (and hence l), and then introduce the quantity

$$u := \lfloor C_2 2^{2\rho l}\rfloor. \tag{5.26}$$

We now combine the previous "arithmetic" dyadic decomposition with a "continuous" dyadic decomposition

$$\Pi_{l_1,\leq -\operatorname{Log} N + l_{(N)}} f = \sum_{-u\leq s_1\leq l_{(N)}} F_N^{u,l_1,s_1},$$

$$\Pi_{l_2,\leq -d\operatorname{Log} N + dl_{(N)}} g = \sum_{-u\leq s_2\leq l_{(N)}} G_N^{u,l_2,s_2},$$

where

$$F_N^{u,l_1,s_1} := \begin{cases} \Pi_{l_1,\leq -\operatorname{Log} N + s_1} f - \Pi_{l_1,\leq -\operatorname{Log} N + s_1 - 1} f & s_1 > -u, \\ \Pi_{l_1,\leq -\operatorname{Log} N - u} f & s_1 = -u \end{cases} \tag{5.27}$$

and

$$G_N^{u,l_2,s_2} := \begin{cases} \Pi_{l_2,\leq d(-\operatorname{Log} N+s_2)} g - \Pi_{l_2,\leq d(-\operatorname{Log} N+s_2-1)} g & s_2 > -u, \\ \Pi_{l_2,\leq d(-\operatorname{Log} N-u)} g & s_2 = -u. \end{cases} \tag{5.28}$$

Informally, $F_N^{u,l_1,-u}$, $G_N^{u,l_2,-u}$ represent the "low (continuous) frequency" components of f, g respectively, whereas $F_N^{u,l_1,s_1}, s_1 > -u$ and $G_N^{u,l_2,s_2}, s_2 > -u$ represent the "high (continuous) frequency" components.

By the triangle inequality, we can bound the left-hand side of (5.25) by

$$\sum_{s_1,s_2\geq -u} \|(\tilde{A}_N(F_N^{u,l_1,s_1}, G_N^{u,l_2,s_2}))_{N\in\mathbb{I}^{l,s_1,s_2}}\|_{\ell^p(\mathbb{Z};\boldsymbol{V}^r)},$$

where $\mathbb{I}^{l,s_1,s_2}$ denotes the index set

$$\mathbb{I}^{l,s_1,s_2} := \{N \in \mathbb{D} : l, s_1, s_2 \leq l_{(N)}\}. \tag{5.29}$$

The expression $\tilde{A}_N(F_N^{u,l_1,s_1}, G_N^{u,l_2,s_2})$ can be viewed as (the scale N component of) a paraproduct of F and G, but centered around a finite number of (arithmetic) frequencies, in contrast to the classical paraproducts that are centered at the frequency origin; also, the paraproduct symbol exhibits some additional oscillation compared to classical paraproducts when s_1, s_2 become large. We shall sometimes distinguish between the "high-high" case $s_1, s_2 > -u$, the "low-high" case $s_2 > s_1 = -u$, the "high-low" case $s_1 > s_2 = -u$, and the "low-low" case $s_1 = s_2 = -u$ of these paraproducts. But for now we can treat all choices of s_1, s_2 in a unified fashion.

By several applications of the triangle inequality, the bound (5.25), and hence Theorem 3.9, now follows from the following variational paraproduct estimates, in which we request an exponential gain in the $p_1 = p_2 = 2$ case and relatively small losses in all other cases:

THEOREM 5.30 (Variational paraproduct estimates). *Let the hypotheses be as in Theorem 3.9, and let the notational conventions be as in this section. Let $l_1, l_2 \in \mathbb{N}$, and define l, u by (5.24) and (5.26) respectively. Let $s_1, s_2 \geq -u$, and then let $F_N := F_N^{u,l_1,s_1}$, $G_N := G_N^{u,l_2,s_2}$, $\mathbb{I} := \mathbb{I}^{l,s_1,s_2}$ be defined respectively by (5.27), (5.28), and (5.29). Then*

$$\begin{aligned} &\|(\tilde{A}_N(F_N, G_N))_{N\in\mathbb{I}}\|_{\ell^p(\mathbb{Z};\boldsymbol{V}^r)} \\ &\lesssim_{C_3} \langle \max(l, s_1, s_2)\rangle^{O(1)} 2^{O(\rho l) - c\max(l,s_1,s_2)\mathbb{1}_{p_1=p_2=2}} \|f\|_{\ell^{p_1}(\mathbb{Z})} \|g\|_{\ell^{p_2}(\mathbb{Z})}. \end{aligned} \tag{5.31}$$

Here the constant c does not depend on ρ; see the discussion below Theorem 5.12.

Indeed, by interpolating (5.31) between the case $(p_1, p_2, p) = (2, 2, 1)$ and the case where (p_1, p_2, p) are close to $(1, \infty, 1)$, $(\infty, 1, 1)$, or (∞, ∞, ∞), we see

that (say)

$$\begin{aligned}\|(\tilde{A}_N(F_N,G_N))_{N\in\mathbb{I}}\|_{\ell^p(\mathbb{Z};\boldsymbol{V}^r)}\\ \lesssim_{C_3} \langle \max(l,s_1,s_2)\rangle^{O(1)} 2^{-10\rho\max(l,s_1,s_2)}\|f\|_{\ell^{p_1}(\mathbb{Z})}\|g\|_{\ell^{p_2}(\mathbb{Z})}.\end{aligned} \tag{5.32}$$

Then using $\langle a\rangle^{O(1)}2^{-10\rho a} \lesssim_{C_3} 2^{-8\rho a}$ for $a \geq 0$ and summing the bound in (5.32) over s_1, s_2, we see that to obtain the (5.25) from Theorem 5.30, it suffices to establish the bound

$$\sum_{s_1,s_2\geq -u} 2^{-8\rho\max(l,s_1,s_2)} \lesssim_{C_3} 2^{-\rho l};$$

bounding

$$2^{-8\rho\max(l,s_1,s_2)} \leq 2^{-4\rho\max(l,s_1)}2^{-4\rho\max(l,s_2)}$$

it suffices to show that

$$\sum_{s_0\geq -u} 2^{-4\rho\max(l,s_0)} \lesssim_{C_3} 2^{-\rho l/2}.$$

But this is clear from the geometric series formula since there are only $O_{C_2}(2^{2\rho l})$ scales s_0 with $-u \leq s_0 \leq l$.

It remains to establish Theorems 5.12 and 5.30. Theorem 5.12 will be established in the next section; the rest of the paper is then devoted to the proof of Theorem 5.30. For now, we use Theorem 5.12 to deal with one case of Theorem 5.30:

PROPOSITION 5.33 (High-high $\ell^2(\mathbb{Z})$ case). *Theorem* 5.30 *holds when* $s_1, s_2 > -u$ *and* $p_1 = p_2 = 2$.

In view of this proposition, for the purposes of proving Theorem 5.30 we may assume that at least one of $s_1 = -u$, $s_2 = -u$, or $(p_1, p_2) \neq (2, 2)$ holds.

Proof. From (2.7), we have

$$\|(\tilde{A}_N(F_N,G_N))_{N\in\mathbb{I}}\|_{\ell^1(\mathbb{Z};\boldsymbol{V}^r)} \lesssim \sum_{N\in\mathbb{I}} \|\tilde{A}_N(F_N,G_N)\|_{\ell^1(\mathbb{Z})}.$$

Observe (using Lemma 5.2 and (5.27)) that for $N \in \mathbb{I}$, $\mathcal{F}_{\mathbb{Z}}F_N$ vanishes on the major arcs $\mathcal{M}_{\leq\max(l_1,s_1)-1,\leq -\operatorname{Log} N+\max(l_1,s_1)-1}$, and hence by Theorem 5.12 we have

$$\|\tilde{A}_N(F_N,G_N)\|_{\ell^1(\mathbb{Z})} \lesssim_{C_1} (2^{-c\max(l_1,s_1)} + \langle \operatorname{Log} N\rangle^{-cC_1})\|F_N\|_{\ell^2(\mathbb{Z})}\|G_N\|_{\ell^2(\mathbb{Z})}.$$

A similar argument gives

$$\|\tilde{A}_N(F_N,G_N)\|_{\ell^1(\mathbb{Z})} \lesssim_{C_1} (2^{-c\max(l_2,s_2)} + \langle \operatorname{Log} N\rangle^{-cC_1})\|F_N\|_{\ell^2(\mathbb{Z})}\|G_N\|_{\ell^2(\mathbb{Z})},$$

and hence on taking geometric means,

$$\|\tilde{A}_N(F_N,G_N)\|_{\ell^1(\mathbb{Z})} \lesssim_{C_1} (2^{-c\max(l,s_1,s_2)} + \langle \operatorname{Log} N\rangle^{-cC_1})\|F_N\|_{\ell^2(\mathbb{Z})}\|G_N\|_{\ell^2(\mathbb{Z})}.$$

From (5.29) we have $\langle \operatorname{Log} N\rangle^{-cC_1} \lesssim_{C_3} 2^{-c\max(l,s_1,s_2)}$, hence

$$\|\tilde{A}_N(F_N,G_N)\|_{\ell^1(\mathbb{Z})} \lesssim_{C_3} 2^{-c\max(l,s_1,s_2)}\|F_N\|_{\ell^2(\mathbb{Z})}\|G_N\|_{\ell^2(\mathbb{Z})}.$$

By the Cauchy–Schwarz inequality, it thus suffices to establish the Bessel-type inequalities

$$\sum_{N\in\mathbb{I}} \|F_N\|_{\ell^2(\mathbb{Z})}^2 \lesssim \|f\|_{\ell^2(\mathbb{Z})}^2$$

and

$$\sum_{N\in\mathbb{I}} \|G_N\|_{\ell^2(\mathbb{Z})}^2 \lesssim \|g\|_{\ell^2(\mathbb{Z})}^2.$$

But this follows from the easily verified pointwise bounds

$$\sum_{N\in\mathbb{I}} |\mathcal{F}_{\mathbb{Z}}F_N(\xi)|^2 \lesssim |\mathcal{F}_{\mathbb{Z}}f(\xi)|^2,$$

$$\sum_{N\in\mathbb{I}} |\mathcal{F}_{\mathbb{Z}}G_N(\xi)|^2 \lesssim |\mathcal{F}_{\mathbb{Z}}g(\xi)|^2$$

and Plancherel's theorem. □

6. Minor arc single scale estimate: applying Peluse–Prendiville theory

In this section we establish Theorem 5.12. The arguments here will be lengthy, but they are not needed elsewhere in this paper.

It will be convenient to exploit duality and work with trilinear forms $\langle \tilde{A}_N(f,g),h\rangle$ instead of bilinear operators $\tilde{A}_N(f,g)$. We use the inner product

$$\langle f,g\rangle := \sum_{x\in\mathbb{Z}} f(x)g(x)$$

on $\boldsymbol{S}(\mathbb{Z})$. (There will be no advantage for us in this bilinear analysis in inserting a complex conjugation into the inner product.) We observe the identities

$$\langle \tilde{A}_N(f,g),h\rangle = \langle \tilde{A}_N^*(h,g),f\rangle = \langle \tilde{A}_N^{**}(f,h),g\rangle \tag{6.1}$$

for $f,g,h\in\boldsymbol{S}(\mathbb{Z})$, where the transpose operators $\tilde{A}_N^*,\tilde{A}_N^{**}$ are the averaging operators

(6.2)

$$\tilde{A}_N^*(h,g)(x) := \tilde{A}_N^{-\mathrm{n},P(\mathrm{n})-\mathrm{n}}(h,g)(x) = \mathbb{E}_{n\in[N]}h(x+n)g(x+n-P(n))\mathbb{1}_{n>N/2}$$

and

$$\begin{aligned}\tilde{A}_N^{**}(f,h)(x) &:= \tilde{A}_N^{\mathrm{n}-P(\mathrm{n}),-P(\mathrm{n})}(f,h)(x)\\ &= \mathbb{E}_{n\in[N]}f(x+P(n)-n)h(x+P(n))\mathbb{1}_{n>N/2}.\end{aligned}$$

In the language of additive combinatorics, the functions $\tilde{A}_N^*(h,g),\tilde{A}_N^{**}(f,h)$ are referred to as *dual functions*.

6.1. *Proof of Theorem 5.12*(i). Our starting point is the following deep inverse theorem of Peluse–Prendiville [74] in the quadratic case $P(\mathrm{n}) = \mathrm{n}^2$ (see also [75] and [77]), and Peluse [73] for general polynomials $P(\mathrm{n})$ of degree $d \geq 2$.

THEOREM 6.3 (Peluse inverse theorem). *Let $N \geq 1$ and $0 < \delta \leq 1$, and let N_0 be a quantity with $N_0 \sim N^d$. Let $f, g, h \in \boldsymbol{S}(\mathbb{Z})$ be supported on $[-N_0, N_0]$ with $\|f\|_{\ell^\infty(\mathbb{Z})}, \|g\|_{\ell^\infty(\mathbb{Z})}, \|h\|_{\ell^\infty(\mathbb{Z})} \leq 1$, obeying the lower bound*

$$|\langle \tilde{A}_N(f,g), h\rangle| \geq \delta N^d. \tag{6.4}$$

Then one of the following holds:

(i) (N not too large). *One has $N \lesssim \delta^{-O(1)}$.*

(ii) (f has major arc structure at scale N). *There exist a positive integer $q \lesssim \delta^{-O(1)}$ and a positive integer $\delta^{O(1)} N \lesssim N' \leq N$ such that*

$$\frac{1}{N^d}\Big|\sum_{x\in\mathbb{Z}} \mathbb{E}_{m\in[N']} f(x+qm)\Big| \gtrsim \delta^{O(1)}.$$

Note from the uncertainty principle (cf. Remark 5.20) that conclusion (ii) of Theorem 6.3 is morally equivalent to asserting that the Fourier transform $\mathcal{F}_{\mathbb{Z}} f$ has a large presence on a major arc set $\mathcal{M}_{\leq l, \leq k, \mathrm{naive}}$ with $2^l \lesssim \delta^{-O(1)}$ and $2^{-k} \lesssim \delta^{-O(1)}/N$. This intuition will be formalized in Proposition 6.6 below.

Proof. We expand out (6.4) as

$$\frac{1}{N^{d+1}}\Big|\sum_{n\in[N]}\sum_{x\in\mathbb{Z}} h(x) f(x-n) g(x-P(n)) \mathbb{1}_{n>N/2}\Big| \geq \delta.$$

By the triangle inequality, we thus have

$$\frac{1}{(N')^{d+1}}\Big|\sum_{n\in[N']}\sum_{x\in\mathbb{Z}} h(x) f(x-n) g(x-P(n))\Big| \gtrsim \delta$$

for either $N' = N$ or $N' = \lfloor N/2 \rfloor$. The claim now follows from [73, Th. 3.3] (after some minor changes of notation) with parameters $(m, q, N, M, P_1, P_2) = (2, 1, N_0, N', \mathrm{n}, P(\mathrm{n}))$. In that theorem, the functions f, g, h were assumed to be supported on $[1, (N')^d]$ rather than $[-N_0, N_0]$, but it is a routine matter to see that the arguments continue to hold with this slightly more general support hypothesis. □

We will now gradually manipulate Theorem 6.3 in a sequence of steps to make it more closely resemble (the contrapositive of) Theorem 5.12(i), until we are able to actually establish that part of the theorem; we will then adapt the argument (focusing on g instead of f) to also establish Theorem 5.12(ii).

The first step is to make the conclusion of Theorem 6.3 more Fourier-analytic in nature. We need a technical calculation:

LEMMA 6.5 (Smooth approximation to $\mathbb{1}_{[a,b]}$). *Let $\psi \in \boldsymbol{S}(\mathbb{R})$ with $\int_{\mathbb{R}} \psi(x)\,dx = 1$. Then for any interval $[a,b] \subset \mathbb{R}$ and any $0 < \varepsilon \le 1$, one has the pointwise bound*

$$\sum_{y \in [a,b] \cap \mathbb{Z}} \varepsilon\psi(\varepsilon(x-y)) - \mathbb{1}_{[a,b]}(x) \lesssim_\psi \varepsilon^{10} + \langle \varepsilon(x-a)\rangle^{-10} + \langle \varepsilon(x-b)\rangle^{-10}$$

for all $x \in \mathbb{Z}$.

Proof. By the triangle inequality, it suffices to show that

$$\sum_{y \in \mathbb{Z}: y \ge a} \varepsilon\psi(\varepsilon(x-y)) - \mathbb{1}_{x \ge a} \lesssim_\psi \varepsilon^{10} + \langle \varepsilon(x-a)\rangle^{-10}$$

since the claim then follows by subtracting this estimate from the analogous estimate for b (adjusting b by an infinitesimal amount if necessary). By translation invariance, we may set $a = 0$. From the Poisson summation formula and the rapid decrease of $\mathcal{F}\psi$, one has

$$\sum_{y \in \mathbb{Z}} \varepsilon\psi(\varepsilon(x-y)) = 1 + O_\psi(\varepsilon^{10}),$$

so by reflection symmetry and the triangle inequality, it suffices to show that

$$\sum_{y \in \mathbb{Z}: y \ge 0} \varepsilon\psi(\varepsilon(x-y)) \lesssim_\psi \langle \varepsilon x\rangle^{-10}$$

when $x < 0$. But this follows from the rapid decrease of ψ. □

PROPOSITION 6.6 (Alternate inverse theorem for f). *Under the hypotheses and notation of Theorem 6.3, there exists a function $F \in \ell^2(\mathbb{Z})$ with*

$$\|F\|_{\ell^\infty(\mathbb{Z})} \lesssim 1; \quad \|F\|_{\ell^1(\mathbb{Z})} \lesssim N^d \tag{6.7}$$

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

$$|\langle f, F\rangle| \gtrsim \delta^{O(1)} N^d. \tag{6.8}$$

Proof. If $N \lesssim \delta^{-O(1)}$, then we can simply take $F = \tilde{A}_N^*(h,g)$ and $a/q = 1/1$ and use (6.1) and (1.6) to conclude. Thus we may assume that $N \ge C_* \delta^{-C_*}$ for a sufficiently large C_*. In particular, by Theorem 6.3, we can find $N', q \in \mathbb{Z}_+$ with $q \lesssim \delta^{-O(1)}$ and $\delta^{O(1)} N \lesssim N' \le N$ such that

$$\sum_{x \in \mathbb{Z}} |\mathbb{E}_{m \in [N']} f(x+qm)| \gtrsim \delta^{O(1)} N^d.$$

Observe that the summand vanishes unless $|x| \le N_0 + O(qN') \lesssim N^d$, thus

$$\sum_{x = O(N^d)} |\mathbb{E}_{m \in [N']} f(x+qm)| \gtrsim \delta^{O(1)} N^d.$$

Now we smooth out the inner average $\mathbb{E}_{m\in[N']}$. Let $0 < \varepsilon \leq 1$ be a parameter to be chosen later. From Lemma 6.5, one has

$$\mathbb{1}_{[N']}(m) = \varepsilon \sum_{m'\in[N']} \mathcal{F}_{\mathbb{R}}^{-1}\eta(\varepsilon(m-m')) + O(\varepsilon^{10} + \langle \varepsilon m\rangle^{-10} + \langle \varepsilon(m-N')\rangle^{-10})$$

for any $m \in \mathbb{Z}$, where η is the cutoff from Section 2.3. Hence from the boundedness of f,

$$\mathbb{E}_{m\in[N']} f(x+qm) = \varepsilon \sum_{m\in\mathbb{Z}} \mathbb{E}_{m'\in[N']}\mathcal{F}_{\mathbb{R}}^{-1}\eta(\varepsilon(m-m'))f(x+qm) + O\left(\varepsilon^{10} + \frac{1}{\varepsilon N'}\right).$$

If we choose $\varepsilon := C\delta^{-C}/N$ for some large C (depending only on η, P) and take C_* large enough depending on C, we conclude that

$$\sum_{x=O(N^d)} \left|\varepsilon \sum_{m\in\mathbb{Z}} \mathbb{E}_{m'\in[N']}\mathcal{F}_{\mathbb{R}}^{-1}\eta(\varepsilon(m-m'))f(x+qm)\right| \gtrsim \delta^{O(1)}N^d.$$

In the latter case, there exists $G \in \ell^\infty(\mathbb{Z})$ supported on $[-O(N^d), O(N^d)]$ with $\|G\|_{\ell^\infty(\mathbb{Z})} \leq 1$, with

$$\left|\sum_{x\in\mathbb{Z}} G(x)\varepsilon \sum_{m\in\mathbb{Z}} \mathbb{E}_{m'\in[N']}\mathcal{F}_{\mathbb{R}}^{-1}\eta(\varepsilon(m-m'))f(x+qm)\right| \gtrsim \delta^{O(1)}N^d.$$

We thus have the claim (6.8) with

$$F(x) := \varepsilon \sum_{m\in\mathbb{Z}} \mathbb{E}_{m'\in[N']}\mathcal{F}_{\mathbb{R}}^{-1}\eta(\varepsilon(m-m'))G(x-qm).$$

From the hypotheses on η, G, we easily verify the bounds (6.7). A routine calculation using the Poisson summation formula reveals the identity

$$\mathcal{F}_{\mathbb{Z}}F(\xi \bmod 1) = \mathcal{F}_{\mathbb{Z}}G(\xi \bmod 1)\mathbb{E}_{m'\in[N']}e(qm'\xi)\sum_{n\in\mathbb{Z}} \mathcal{F}_{\mathbb{R}}\mathcal{F}_{\mathbb{R}}^{-1}\eta\left(\frac{q\xi-n}{\varepsilon}\right)$$

for any $\xi \in \mathbb{R}$, which, in particular, implies from the support of η that $\mathcal{F}_{\mathbb{Z}}F$ is supported in the set

$$\pi\left(\left[-\frac{\varepsilon}{q}, \frac{\varepsilon}{q}\right] \times \left(\frac{1}{q}\mathbb{Z}/\mathbb{Z}\right)\right).$$

By applying suitable Fourier multiplier operators, one can then decompose $F = \sum_{a\in[q]} F_a$, where each F_a obeys essentially the same bounds (6.7) as F and is supported in the $\frac{\varepsilon}{q}$-neighborhood of $\frac{a}{q}$ mod 1. The claim now follows from the pigeonhole principle and the bounds on ε, q. □

We now dualize the above proposition using the Hahn–Banach theorem to obtain control on dual functions $\tilde{A}^*(h, g)$. Specifically, we shall use the following lemma.

LEMMA 6.9 (Application of Hahn–Banach). *Let $A, B > 0$, and let G be an element of $\ell^2(\mathbb{Z})$. Let Φ be a family of vectors in $\ell^2(\mathbb{Z})$, and assume the following inverse theorem: whenever $f \in \ell^2(\mathbb{Z})$ is such that $\|f\|_{\ell^\infty(\mathbb{Z})} \le 1$ and $|\langle f, G\rangle| > A$, then $|\langle f, \phi\rangle| > B$ for some $\phi \in \Phi$. Then G lies in the closed convex hull of*

$$V = \{\lambda\phi \in \ell^2(\mathbb{Z}) : \phi \in \Phi,\ |\lambda| \le A/B\} \cup \{h \in \ell^2(\mathbb{Z}) : \|h\|_{\ell^1(\mathbb{Z})} \le A\}.$$

Proof. Observe that the set $\overline{\mathrm{conv} V}^{\|\cdot\|_{\ell^2(\mathbb{Z})}}$ is balanced. Therefore, if the claim of Lemma 6.9 failed, then from the Hahn–Banach theorem and the Riesz representation theorem, there exists $f \in \ell^2(\mathbb{Z})$ such that $\mathrm{Re}\langle f, G\rangle > A$, but $\mathrm{Re}\langle f, h\rangle \le A$ for all $h \in V$. In particular, this gives $|\langle f, h\rangle| \le A$ for all $h \in V$, which implies that

$$|\langle f, \phi\rangle| \le B$$

for all $\phi \in \Phi$, and that

$$\|f\|_{\ell^\infty(\mathbb{Z})} = \sup_{\|h\|_{\ell^1(\mathbb{Z})} \le 1} |\langle f, h\rangle| \le 1,$$

contradicting the hypothesis. This completes the proof of the lemma. □

COROLLARY 6.10 (Structure of dual function, I). *Let $N \ge 1$ and $N_0 \sim N^d$. Let $g, h \in \boldsymbol{S}(\mathbb{Z})$ be supported on $[-N_0, N_0]$ with $\|g\|_{\ell^\infty(\mathbb{Z})}, \|h\|_{\ell^\infty(\mathbb{Z})} \le 1$, and let $0 < \delta \le 1$. Then there exists a decomposition*

$$\tilde{A}_N^*(h, g) = \sum_{\alpha \in \mathbb{Q}/\mathbb{Z} : \mathrm{h}_{\mathrm{naive}}(\alpha) \lesssim \delta^{-O(1)}} F_\alpha + E_1 + E_2, \tag{6.11}$$

where each $F_\alpha \in \ell^2(\mathbb{Z})$ has Fourier transform supported in the $O(\delta^{-O(1)}/N)$-neighborhood of α and obeys the bounds

$$\|F_\alpha\|_{\ell^\infty(\mathbb{Z})} \lesssim \delta^{-O(1)} \quad \text{and} \quad \|F_\alpha\|_{\ell^1(\mathbb{Z})} \lesssim \delta^{-O(1)} N^d, \tag{6.12}$$

and the error terms $E_1 \in \ell^1(\mathbb{Z})$ and $E_2 \in \ell^2(\mathbb{Z})$ obey the bounds

$$\|E_1\|_{\ell^1(\mathbb{Z})} \le \delta N^d \quad \text{and} \quad \|E_2\|_{\ell^2(\mathbb{Z})} \le \delta. \tag{6.13}$$

For similar applications of the Hahn–Banach theorem to analyze the structure of dual functions in additive combinatorics, see [45, pp. 221], [38, Th. 3.8].

Proof. There exists $f \in \ell^\infty(\mathbb{Z})$ with $\|f\|_{\ell^\infty(\mathbb{Z})} \le 1$ such that

$$|\langle f, \tilde{A}_N^*(h, g)\rangle| > \delta N^d.$$

Applying Proposition 6.6, we obtain

$$|\langle f, F\rangle| \gtrsim \delta^{O(1)} N^d$$

for some function $F \in \ell^2(\mathbb{Z})$ obeying the properties of Proposition 6.6. Invoking Lemma 6.9 with $A = \delta N^d/2$ and $B \sim \delta^{O(1)} N^d$ and the set

$$\Phi = \{\phi_\alpha \in \ell^2(\mathbb{Z}) : \alpha \in \mathbb{Q}/\mathbb{Z}; \mathrm{h}_{\mathrm{naive}}(\alpha) \lesssim \delta^{-O(1)}\},$$

we obtain a decomposition

$$\tilde{A}_N^*(h,g) = \sum_{j=1}^{\infty} c_j \phi_j + E_1 + E_2, \tag{6.14}$$

with the following properties:

(i) for each $j \in \mathbb{Z}_+$, we have that $\phi_j = \lambda_j \phi_{\alpha_j}$ for some $\phi_{\alpha_j} \in \Phi$ and $\lambda_j \in \mathbb{C}$ such that $|\lambda_j| \lesssim \delta^{-O(1)}$;
(ii) the coefficients c_j are non-negative with $\sum_{j=1}^{\infty} c_j \leq 1$, and all but finitely c_j vanish;
(iii) the error term $E_1 \in \ell^1(\mathbb{Z})$ satisfies $\|E_1\|_{\ell^1(\mathbb{Z})} \leq \delta N^d$;
(iv) the error term $E_2 \in \ell^2(\mathbb{Z})$ satisfies $\|E_2\|_{\ell^2(\mathbb{Z})} \leq \delta$.

The latter error term arises as a consequence of the fact that one is working with the closed convex hull instead of the convex hull. In fact, its $\ell^2(\mathbb{Z})$ norm can be made arbitrarily small, but δ will suffice for our purposes. Grouping together terms associated to each arithmetic frequency α in (6.14) and using the triangle inequality, we obtain the desired decomposition from (6.11) that satisfies (6.12) and (6.13). □

Corollary 6.10 is not directly suitable for our applications for three reasons: firstly, E_1 is controlled in $\ell^1(\mathbb{Z})$ rather than in $\ell^2(\mathbb{Z})$; secondly, g is required to be controlled in $\ell^\infty(\mathbb{Z})$ rather than in $\ell^2(\mathbb{Z})$; and thirdly the support of g is restricted to an interval. Using the Ionescu–Wainger projections, we now address the first issue, at the cost of worsening the control of the structured component of the decomposition (6.11), and also requiring δ to not be too small.

PROPOSITION 6.15 (Structure of dual function, II). *If $N, N_0 \in \mathbb{Z}_+$ with $N_0 \sim N^d$ and $l \in \mathbb{N}$, with*

$$\operatorname{Log} N \geq C_\rho 2^{\rho l} \tag{6.16}$$

for a sufficiently large constant C_ρ depending on ρ, one has the estimate

$$\|(1 - \Pi_{\leq l, \leq -\operatorname{Log} N + l}) \tilde{A}_N^*(h,g)\|_{\ell^2(\mathbb{Z})} \lesssim_{C_1} 2^{-cl} N^{d/2} \|h\|_{\ell^\infty(\mathbb{Z})} \|g\|_{\ell^\infty(\mathbb{Z})},$$

whenever $g, h \in \boldsymbol{S}(\mathbb{Z})$ are supported on $[-N_0, N_0]$.

Proof. We can assume N is sufficiently large depending on C_1, as the claim follows from (1.6) otherwise. We may also normalize $\|g\|_{\ell^\infty(\mathbb{Z})} = \|h\|_{\ell^\infty(\mathbb{Z})} = 1$, so our task is now to show that

$$\|(1 - \Pi_{\leq l, \leq -\operatorname{Log} N + l}) \tilde{A}_N^*(h,g)\|_{\ell^2(\mathbb{Z})} \lesssim_{C_1} 2^{-cl} N^{d/2}$$

for some $c > 0$ depending on P.

We apply Corollary 6.10 with $\delta = 2^{-c'l}$ for a sufficiently small $c' > 0$ depending only on P. Because of (6.16) and the hypothesis that N is large,

we see from (5.8) that $(l, -\operatorname{Log} N + l)$ has good major arcs. By choice of δ and the Fourier support of F_α, we have from Lemma 5.17 that

$$(1 - \Pi_{\le l, \le -\operatorname{Log} N + l}) F_\alpha = 0$$

for all F_α in the decomposition (6.11), and hence

$$\begin{aligned}(1 - \Pi_{\le l, \le -\operatorname{Log} N + l}) \tilde{A}_N^*(h, g) = {} & (1 - \Pi_{\le l, \le -\operatorname{Log} N + l}) E_1 \\ & + (1 - \Pi_{\le l, \le -\operatorname{Log} N + l}) E_2.\end{aligned} \tag{6.17}$$

Since

$$\|(1 - \Pi_{\le l, \le -\operatorname{Log} N + l}) E_2\|_{\ell^2(\mathbb{Z})} \lesssim \delta,$$

it suffices to show that

$$\|(1 - \Pi_{\le l, \le -\operatorname{Log} N + l}) E_1\|_{\ell^2(\mathbb{Z})} \lesssim \langle l \rangle \delta^{1/4} N^{d/2}, \tag{6.18}$$

which will give the claim by the choice of δ. We now establish (6.18).

The function $\tilde{A}_N^*(h, g)$ is bounded in $\ell^\infty(\mathbb{Z})$ norm by $O(1)$. From (6.11) and the triangle inequality, we thus have

$$\|E_1\|_{\ell^\infty(\mathbb{Z})} \lesssim \delta^{-O(1)},$$

since $E_2 \in \ell^q(\mathbb{Z})$ for any $2 \le q \le \infty$ and $\|E_2\|_{\ell^q(\mathbb{Z})} \le \|E_2\|_{\ell^2(\mathbb{Z})} \le \delta$, so by interpolation with (6.13), we have

$$\|E_1\|_{\ell^p(\mathbb{Z})} \lesssim \delta^{1/2} N^{d/p}$$

for some absolute constant $1 < p < 2$ that is sufficiently close to 1. By the latter bound and Lemma 5.17, we conclude that

$$\|(1 - \Pi_{\le l, \le -\operatorname{Log} N + l}) E_1\|_{\ell^p(\mathbb{Z})} \lesssim_p \langle l \rangle \delta^{1/2} N^{d/p}. \tag{6.19}$$

Furthermore, as $\tilde{A}_N^*(h, g)$ is bounded by $O(1)$ and supported on $[-N_0, N_0]$ with $N_0 \simeq N^d$, we have

$$\|A_N^*(h, g)\|_{\ell^{p'}(\mathbb{Z})} \lesssim_p N^{d/p'}.$$

Thus by Lemma 5.17 again,

$$\|(1 - \Pi_{\le l, \le -\operatorname{Log} N + l}) \tilde{A}_N^*(h, g)\|_{\ell^{p'}(\mathbb{Z})} \lesssim_{p, C_1} \langle l \rangle N^{d/p'},$$

and since $\|E_2\|_{\ell^{p'}(\mathbb{Z})} \lesssim \delta$, we also have

$$\|(1 - \Pi_{\le l, \le -\operatorname{Log} N + l}) E_2\|_{\ell^{p'}(\mathbb{Z})} \lesssim_{p, C_1} \langle l \rangle \delta.$$

Using these two bounds, the triangle inequality and (6.17), we may write

$$\|(1 - \Pi_{\le l, \le -\operatorname{Log} N + l}) E_1\|_{\ell^{p'}(\mathbb{Z})} \lesssim_{p, C_1} \langle l \rangle N^{d/p'}. \tag{6.20}$$

Interpolating (6.19) and (6.20), we obtain (6.18), and the proof is completed. □

We now address the second issue, namely that of relaxing the $\ell^\infty(\mathbb{Z})$ control on g to $\ell^2(\mathbb{Z})$ control. The main tool for this is the following recent $\ell^p(\mathbb{Z})$ improving estimate for linear polynomial averages.

PROPOSITION 6.21 ($\ell^p(\mathbb{Z})$-improving). *Let $Q(\mathrm{n}) \in \mathbb{Z}[\mathrm{n}]$ be of degree $d \geq 2$. Then for every*

$$2 \geq p > \begin{cases} 2 - \frac{4}{d^2+d+3} & \text{if } d \geq 3, \\ 2 - \frac{2}{3} & \text{if } d = 2, \end{cases}$$

one has the bound

$$\|A_N^{Q(\mathrm{n})} f\|_{\ell^2(\mathbb{Z})} \lesssim_{p,Q} N^{d(\frac{1}{2}-\frac{1}{p})} \|f\|_{\ell^p(\mathbb{Z})}$$

for all $N \geq 1$ and $f \in \ell^p(\mathbb{Z})$.

Proof. This follows from the work of Han–Kovač–Lacey–Madrid–Yang [41]. Indeed, the $d = 2$ case is contained[8] in [41, Th. 1.6], and the $d \geq 3$ case is contained in [41, Th. 1.9], after specializing these theorems to the $p = 2$ case and performing some routine algebra. Note that [41, Conj. 1.5] predicts that the range of p can be lowered to $p > 2 - \frac{2}{d+1}$ for any value of d, but this is currently only known for $d = 2$. For our purposes, any exponent p less than 2 would be sufficient for applications. □

We can now relax the $\ell^\infty(\mathbb{Z})$ control on g to $\ell^2(\mathbb{Z})$ control:

COROLLARY 6.22 (Structure of dual function, III). *Under the notation and hypotheses of Proposition 6.15, one has*

$$\|(1 - \Pi_{\leq l, \leq -\operatorname{Log} N + l}) \tilde{A}_N^*(h,g)\|_{\ell^2(\mathbb{Z})} \lesssim_{C_1} 2^{-cl} \|h\|_{\ell^\infty(\mathbb{Z})} \|g\|_{\ell^2(\mathbb{Z})}, \tag{6.23}$$

whenever $g, h \in \boldsymbol{S}(\mathbb{Z})$ are supported on $[-N_0, N_0]$.

Proof. From Proposition 6.15, we already have the bound

$$\|(1 - \Pi_{\leq l, \leq -\operatorname{Log} N + l}) \tilde{A}_N^*(h,g)\|_{\ell^2(\mathbb{Z})} \lesssim_{C_1} N^{d/2} 2^{-cl} \|h\|_{\ell^\infty(\mathbb{Z})} \|g\|_{\ell^\infty(\mathbb{Z})}.$$

On the other hand, from (6.2) and the triangle inequality, we have the pointwise bound

$$\tilde{A}_N^*(h,g)(x) \lesssim \|h\|_{\ell^\infty(\mathbb{Z})} A_N^{\mathrm{n}-P(\mathrm{n})} |g|(x).$$

[8] Strictly speaking, this theorem requires all the coefficients of the quadratic polynomial Q to be non-negative. However, by applying a reflection $x \mapsto -x$ one can assume without loss of generality that the quadratic coefficient of Q is positive, and then applying a translation $n \mapsto n+c$ for some large positive integer c (noting the pointwise bound $A_N^{Q(\mathrm{n})} f \lesssim_c A_N^{Q(\mathrm{n}+c)} |f|$) one can then deduce the case of general Q from the non-negative coefficient case (perhaps at the risk of worsening the dependence of constants on Q). See also [20] for another treatment of the (monomial) quadratic case and an extension to higher dimensions.

Hence by Lemma 5.17(i) and Proposition 6.21 applied with $Q(\mathrm{n}) = \mathrm{n} - P(\mathrm{n})$, we have

$$\|(1 - \Pi_{\le l, \le -\operatorname{Log} N + l})\tilde{A}_N^*(h,g)\|_{\ell^2(\mathbb{Z})} \lesssim_{C_1,p} N^{d(\frac{1}{2}-\frac{1}{p})}\|h\|_{\ell^\infty(\mathbb{Z})}\|g\|_{\ell^p(\mathbb{Z})}$$

for any $2 - \frac{4}{d^2+d+3} < p \le 2$. The claim now follows from interpolation. □

Now we use the off-diagonal decay estimate (5.19) to remove the support condition:

COROLLARY 6.24 (Structure of dual function, IV). *Under the notation and hypotheses of Proposition 6.15, one has (6.23) whenever $g \in \ell^2(\mathbb{Z})$ and $h \in \ell^\infty(\mathbb{Z})$.*

Proof. If g is supported on an interval I of length N^d, then we may restrict h to an $O(N^d)$-neighborhood of I without affecting the average $\tilde{A}_N^*(h,g)$. From Corollary 6.22 and translation invariance, we then conclude that (6.23) holds in this case.

Now we handle the case when g is not supported in such an interval. We may normalize $\|h\|_{\ell^\infty(\mathbb{Z})} = 1$. We can split $g = \sum_{I\in\mathcal{I}} g\mathbb{1}_I$ where I ranges over a partition $\mathcal{I}$ of $\mathbb{R}$ into intervals I of length N^d. Then by the preceding discussion, the local dual function $D_I := \tilde{A}_N^*(h, g\mathbb{1}_I)$ obeys the bound

$$\|(1 - \Pi_{\le l, \le -\operatorname{Log} N + l})D_I\|_{\ell^2(\mathbb{Z})} \lesssim_{C_1} 2^{-cl}\|g\|_{\ell^2(I)} \tag{6.25}$$

for each interval I, and we wish to establish

$$\Big\|\sum_{I\in\mathcal{I}}(1 - \Pi_{\le l, \le -\operatorname{Log} N + l})D_I\Big\|_{\ell^2(\mathbb{Z})} \lesssim_{C_1} 2^{-cl}\|g\|_{\ell^2(\mathbb{Z})}.$$

(Recall c is allowed to vary from line to line.) By squaring and applying Schur's test, it suffices to obtain the decay bound

$$\langle (1 - \Pi_{\le l, \le -\operatorname{Log} N + l})D_I, (1 - \Pi_{\le l, \le -\operatorname{Log} N + l})D_J\rangle \lesssim_{C_1} 2^{-cl}\left\langle \frac{\operatorname{dist}(I,J)}{N^d}\right\rangle^{-2}\|g\|_{\ell^2(I)}\|g\|_{\ell^2(J)}$$

for all intervals I, J of length N^d. From Cauchy–Schwarz and (6.25), we already have

$$\langle (1 - \Pi_{\le l, \le -\operatorname{Log} N + l})D_I, (1 - \Pi_{\le l, \le -\operatorname{Log} N + l})D_J\rangle \lesssim_{C_1} 2^{-cl}\|g\|_{\ell^2(I)}\|g\|_{\ell^2(J)}.$$

On the other hand, $\tilde{A}_N^*(h, g\mathbb{1}_I)$ is supported in a $O(N^d)$-neighborhood of I, and similarly for $\tilde{A}_N^*(h, g\mathbb{1}_J)$; also, $2^{\operatorname{Log} N - l} \lesssim N^d$. From Lemma 5.17(i) and

Cauchy–Schwarz followed by (1.6), we thus have

$$\begin{aligned}\langle (1-\Pi_{\leq l,\leq -\operatorname{Log} N+l})D_I, (1-\Pi_{\leq l,\leq -\operatorname{Log} N+l})D_J\rangle \\ \lesssim_{C_1} \langle l\rangle^{O(1)} \left\langle \frac{\operatorname{dist}(I,J)}{N^d}\right\rangle^{-10} \|D_I\|_{\ell^2(\mathbb{Z})}\|D_J\|_{\ell^2(\mathbb{Z})} \\ \lesssim_{C_1} \langle l\rangle^{O(1)} \left\langle \frac{\operatorname{dist}(I,J)}{N^d}\right\rangle^{-10} \|g\|_{\ell^2(I)}\|g\|_{\ell^2(J)}.\end{aligned}$$

Taking geometric means of the two estimates, we obtain the claim. □

We may now prove Theorem 5.12(i). We may assume that l, N are sufficiently large depending on C_1, since the claim follows from (1.6) otherwise. It suffices to prove this claim under the additional hypothesis (6.16) (which one can view as an upper bound on l in terms of N), since for larger values of l, the hypothesis (i) becomes stronger and the conclusion (5.13) is essentially unchanged. By duality, it now suffices to establish the bound

$$\langle \tilde{A}_N(f,g), h\rangle \lesssim_{C_1} 2^{-cl}\|f\|_{\ell^2(\mathbb{Z})}\|g\|_{\ell^2(\mathbb{Z})}\|h\|_{\ell^\infty(\mathbb{Z})}$$

for any $f \in \ell^2(\mathbb{Z}), g \in \ell^2(\mathbb{Z}), h \in \ell^\infty(\mathbb{Z})$ obeying the hypothesis in Theorem 5.12(i). From (6.1) and Lemma 5.17, we can write the left-hand side as

$$\langle (1-\Pi_{\leq l,\leq -\operatorname{Log} N+l})\tilde{A}^*_N(h,g), f\rangle.$$

The claim now follows from Corollary 6.24 and Cauchy–Schwarz.

6.2. *Proof of Theorem 5.12(ii).* Now we prove Theorem 5.12(ii). This will follow from a similar argument used to prove Theorem 5.12(i), once we establish an analogue of Proposition 6.6 for the function g (with the denominator N in the intervals replaced with N^d). Such a result was obtained very recently in the quadratic case $P = \mathrm{n}^2$ by Peluse and Prendiville [75, Cor. 1.4]. The arguments there likely extend to cover all nonlinear polynomials P. We give a derivation here that is self-contained (except for Theorem 6.3, which is used as a "black box"), inspired by some earlier unpublished notes in this direction by Peluse and Prendiville (private communication).

PROPOSITION 6.26 (Alternate inverse theorem for g). *Under the hypotheses and notation of Theorem 6.3, there exists a function $G \in \ell^2(\mathbb{Z})$ with*

$$\|G\|_{\ell^\infty(\mathbb{Z})} \lesssim 1; \quad \|G\|_{\ell^1(\mathbb{Z})} \lesssim N^d$$

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

$$|\langle g, G\rangle| \gtrsim \delta^{O(1)} N^d. \tag{6.27}$$

Proof. As in the proof of Proposition 6.6 we may assume that $N \geq C\delta^{-C}$ for some large constant C, as the claim is trivial otherwise. From (6.4) and (6.1), we have

$$|\langle f, \tilde{A}_N^*(h,g)\rangle| \geq \delta N^d.$$

Since $\|f\|_{\ell^2(\mathbb{Z})} \lesssim N^{d/2}$, we conclude using the Cauchy–Schwarz inequality that

$$|\langle \tilde{A}_N^*(h,g), \tilde{A}_N^*(h,g)\rangle| \gtrsim \delta^2 N^d.$$

We apply Corollary 6.10 to the second factor $\tilde{A}_N^*(h,g)$, with δ replaced by $c_0\delta^2$ for some small constant $c_0 > 0$, to obtain a decomposition

$$\tilde{A}_N^*(h,g) = \sum_{\alpha \in \mathbb{Q}/\mathbb{Z}: \mathrm{h}_{\mathrm{naive}}(\alpha) \lesssim_{c_0} \delta^{-O(1)}} F_\alpha + E_1 + E_2,$$

where each $F_\alpha \in \ell^2(\mathbb{Z})$ has Fourier support in the $1/M$-neighborhood of α with $M \sim_{c_0} \delta^{O(1)}N$ and obeys the bounds

$$\|F_\alpha\|_{\ell^\infty(\mathbb{Z})} \lesssim_{c_0} \delta^{-O(1)} \quad \text{and} \quad \|F_\alpha\|_{\ell^1(\mathbb{Z})} \lesssim_{c_0} \delta^{-O(1)} N^d, \tag{6.28}$$

and the error terms $E_1 \in \ell^1(\mathbb{Z})$ and $E_2 \in \ell^2(\mathbb{Z})$ obey the bounds

$$\|E_1\|_{\ell^1(\mathbb{Z})} \leq c_0\delta^2 N^d \quad \text{and} \quad \|E_2\|_{\ell^2(\mathbb{Z})} \leq c_0\delta^2.$$

From (1.6) and Hölder's inequality one has

$$|\langle \tilde{A}_N^*(h,g), E_1\rangle| + |\langle \tilde{A}_N^*(h,g), E_2\rangle| \lesssim c_0\delta^2 N^d.$$

Hence if c_0 is small enough, we conclude from the triangle inequality and pigeonhole principle that

$$|\langle \tilde{A}_N^*(h,g), F_\alpha\rangle| \gtrsim \delta^{O(1)} N^d$$

for some $\alpha \in \mathbb{Q}/\mathbb{Z}$ of naive height $O_{c_0}(\delta^{-O(1)})$. Henceforth we suppress the dependence of constants on c_0. By (6.1) again, we conclude that

$$\Big|\sum_{x\in\mathbb{Z}} \mathbb{E}_{n\in[N]} h(x) F_\alpha(x-n) g(x-P(n))\Big| \gtrsim \delta^{O(1)} N^d.$$

From the Fourier support of F_α, we have the reproducing formula

$$F_\alpha(x) = \frac{2}{M} \sum_{m\in\mathbb{Z}} F_\alpha(x-m) e(-\alpha m) \mathcal{F}_{\mathbb{R}}^{-1}\eta(2m/M),$$

where η was defined in Section 2.3. Thus

$$\Big|\sum_{x\in\mathbb{Z}} \mathbb{E}_{n\in[N]} \sum_{m\in\mathbb{Z}} h(x) F_\alpha(x-m-n) e(-\alpha m) g(x-P(n)) \mathcal{F}_{\mathbb{R}}^{-1}\eta(2m/M)\Big| \gtrsim \delta^{O(1)} N^{d+1}.$$

Making the change of variables $s = m + n$, the left-hand side can be rewritten as

$$\Big|\sum_{x\in\mathbb{Z}}\sum_{s\in\mathbb{Z}} h(x)F_\alpha(x-s)e(-\alpha s)\mathbb{E}_{n\in[N]}e(\alpha n)g(x-P(n))\mathcal{F}_{\mathbb{R}}^{-1}\eta(2(s-n)/M)\Big|.$$

By the rapid decay of $\mathcal{F}_{\mathbb{R}}^{-1}\eta$, the inner sum can be restricted to $s = O(N)$. Thus by the pigeonhole principle, there exists $s = O(N)$ such that

$$\Big|\sum_{x\in\mathbb{Z}} h(x)F_\alpha(x-s)e(-\alpha s)\mathbb{E}_{n\in[N]}e(\alpha n)g(x-P(n))\mathcal{F}_{\mathbb{R}}^{-1}\eta(2(s-n)/M)\Big| \gtrsim \delta^{O(1)}N^d.$$

From (6.28) and the boundedness of h, one has

$$\sum_{x\in\mathbb{Z}} |h(x)F_\alpha(x-s)e(-\alpha s)|^2 \lesssim \delta^{-O(1)}N^d.$$

Hence by the Cauchy–Schwarz inequality,

$$\sum_{x\in\mathbb{Z}} \big|\mathbb{E}_{n\in[N]}e(\alpha n)g(x-P(n))\mathcal{F}_{\mathbb{R}}^{-1}\eta(2(s-n)/M)\big|^2 \gtrsim \delta^{O(1)}N^d.$$

By Plancherel's theorem, we can write the left-hand side as

$$\int_{\mathbb{T}} |\mathcal{F}_{\mathbb{Z}}g(\xi)|^2|S_N(\xi)|^2 d\xi,$$

where S_N is the normalized exponential sum

$$S_N(\xi) := \mathbb{E}_{n\in[N]}e(\alpha n)e(\xi P(n))\mathcal{F}_{\mathbb{R}}^{-1}\eta(2(s-n)/M).$$

By another appeal to Plancherel's theorem, one has

$$\int_{\mathbb{T}} |\mathcal{F}_{\mathbb{Z}}g(\xi)|^2 d\xi = \|g\|_{\ell^2(\mathbb{Z})}^2 \lesssim N^d,$$

thus one must have

$$\int_{\Omega} |\mathcal{F}_{\mathbb{Z}}g(\xi)|^2|S_N(\xi)|^2 d\xi \gtrsim \delta^{O(1)}N^d$$

for a set $\Omega \subseteq \mathbb{T}$ of the form

$$\Omega := \{\xi \in \mathbb{T} \colon |S_N(\xi)| \gtrsim \delta^{O(1)}\}.$$

By the inverse form of Weyl's exponential sum estimate (see the argument as in [40, Lemma A.11, pp. 1922]), we obtain

$$\Omega \subseteq \pi([-1/M', 1/M'] \times \{\alpha' \in \mathbb{Q}/\mathbb{Z} : \mathrm{h}_{\mathrm{naive}}(\alpha') \lesssim \delta^{-O(1)}\})$$

for some $M' \sim \delta^{O(1)}N^d$. By the pigeonhole principle, we may therefore find $\alpha' \in \mathbb{Q}/\mathbb{Z}$ of naive height $O(\delta^{-O(1)})$ such that

$$\int_{\alpha'-1/M'}^{\alpha'+1/M'} |\mathcal{F}_{\mathbb{Z}}g(\xi \bmod 1)|^2 d\xi \gtrsim \delta^{O(1)}N^d.$$

By Plancherel's theorem, this implies that

$$\sum_{x\in\mathbb{Z}}\left|\frac{1}{M'}\sum_{m\in\mathbb{Z}}g(x-m)e(-\alpha' m)\mathcal{F}_{\mathbb{R}}^{-1}\eta\left(\frac{2m}{M'}\right)\right|^2 \gtrsim \delta^{O(1)}N^d$$

so that (6.27) holds with

$$G(x) := \frac{1}{(M')^2}\sum_{m\in\mathbb{Z}}\sum_{m'\in\mathbb{Z}}g(x-m+m')e(-\alpha'(m-m'))\mathcal{F}_{\mathbb{R}}^{-1}\eta\left(\frac{2m}{M'}\right)\mathcal{F}_{\mathbb{R}}^{-1}\eta\left(\frac{2m'}{M'}\right).$$

A routine calculation reveals that G has Fourier support in the $2/M'$-neighborhood of α' and obeys the bounds

$$\|G\|_{\ell^\infty(\mathbb{Z})} \lesssim 1 \quad \text{and} \quad \|G\|_{\ell^1(\mathbb{Z})} \lesssim N^d;$$

the claim follows. □

We can now repeat all of the previous arguments with the role of f now played by g, and with the spatial scale N replaced by N^d. For the convenience of the reader, we state the analogous key propositions. Repeating the Hahn–Banach proof of Corollary 6.10, but using Proposition 6.26 in place of Proposition 6.6, we conclude that

COROLLARY 6.29 (Structure of second dual function, I). *Let the notation and hypotheses be as in Corollary 6.10. Then there exists a decomposition*

$$\tilde{A}_N^{**}(f,h) = \sum_{\alpha\in\mathbb{Q}/\mathbb{Z}:\mathrm{h}_{\mathrm{naive}}(\alpha)\lesssim\delta^{-O(1)}} F_\alpha + E_1 + E_2,$$

where each $F_\alpha \in \ell^2(\mathbb{Z})$ has Fourier transform supported in the $O(\delta^{-O(1)}/N^d)$-neighborhood of α and obeys the bounds from (6.12), and the error terms $E_1 \in \ell^1(\mathbb{Z})$ and $E_2 \in \ell^2(\mathbb{Z})$ obey the bounds from (6.13).

Repeating the proof of Proposition 6.15, we conclude that

PROPOSITION 6.30 (Structure of second dual function, II). *Let the notation and hypotheses be as in Proposition 6.15. Then*

$$\|(1-\Pi_{\le l,\le -d\log N+dl})\tilde{A}_N^{**}(f,h)\|_{\ell^2(\mathbb{Z})} \lesssim_{C_1} 2^{-cl}N^{d/2}\|f\|_{\ell^\infty(\mathbb{Z})}\|h\|_{\ell^\infty(\mathbb{Z})},$$

whenever $f,h \in \boldsymbol{S}(\mathbb{Z})$ are supported on $[-N_0,N_0]$.

Repeating the L^p-improving argument used to prove Corollary 6.22, we conclude that

COROLLARY 6.31 (Structure of second dual function, III). *Under the notation and hypotheses of Proposition 6.15, one has*

$$\|(1-\Pi_{\le l,\le -d\log N+dl})\tilde{A}_N^{**}(f,h)\|_{\ell^2(\mathbb{Z})} \lesssim_{C_1} 2^{-cl}\|f\|_{\ell^2(\mathbb{Z})}\|h\|_{\ell^\infty(\mathbb{Z})}, \tag{6.32}$$

whenever $f,h \in \boldsymbol{S}(\mathbb{Z})$ are supported on $[-N_0,N_0]$.

Finally, we repeat the off-diagonal estimate argument used to prove Corollary 6.24 to conclude that

COROLLARY 6.33 (Structure of second dual function, IV). *Under the notation and hypotheses of Proposition 6.15, one has (6.32) whenever $f \in \ell^2(\mathbb{Z})$ and $h \in \ell^\infty(\mathbb{Z})$.*

Theorem 5.12(ii) now follows by repeating the proof of Theorem 5.12(i).

7. Approximation by model operators

To conclude the proof of Theorem 1.17, we need to establish Theorem 5.30. Let $l_1, l_2 \in \mathbb{N}$, and define l, u by (5.24) and (5.26) respectively. Fix $s_1, s_2 \geq -u$. In view of Proposition 5.33 we may assume that at least one of $s_1 = -u$, $s_2 = -u$, $(p_1, p_2) \neq (2,2)$ holds. It will be convenient to adopt the following definition. If $\mathbb{G} = \mathbb{Z}$ or $\mathbb{G} = \mathbb{A}_{\mathbb{Z}}$, we declare a tuple $(H_N)_{N \in \mathbb{I}'}$ of functions $H_N \in L^p(\mathbb{G})$ to be *acceptable* if one has the estimate

$$\begin{aligned}&\|(H_N)_{N\in\mathbb{I}'}\|_{L^p(\mathbb{G};\boldsymbol{V}^r)}\\ &\quad\lesssim_{C_3} \langle \max(l, s_1, s_2)\rangle^{O(1)} 2^{O(\rho l) - c\max(l,s_1,s_2)\mathbb{1}_{p_1=p_2=2}} \|f\|_{\ell^{p_1}(\mathbb{Z})}\|g\|_{\ell^{p_2}(\mathbb{Z})}.\end{aligned}$$

Our task is thus to show that the tuple

$$(\tilde{A}_N(F_N, G_N))_{N\in\mathbb{I}}$$

is acceptable.

The main difficulty here is that the scale parameter N affects the average $\tilde{A}_N(F_N, G_N)$ in three different ways, as the functions F_N, G_N both separately depend on N, and the averaging operator $\tilde{A}_N$ also depends on N. The strategy will be to perform Fourier-analytic manipulations (on the adelic frequency space $\mathbb{R} \times \mathbb{Q}/\mathbb{Z}$) to approximate this expression $\tilde{A}_N(F_N, G_N)$ by linear combinations of simpler "model expressions" $A(\tilde{F}_N, \tilde{G}_N)$, where the functions $\tilde{F}_N, \tilde{G}_N$ still depend on N, but the bilinear averaging operator A is independent of N. In such a setting we will be able to use general arguments (e.g., Rademacher–Menshov type inequalities) to control the variational norms of the bilinear expressions $A(\tilde{F}_N, \tilde{G}_N)$ by variational norms of the two linear expressions $\tilde{F}_N, \tilde{G}_N$ separately. These in turn can be controlled by a number of tools, such as the vector-valued Ionescu–Wainger multiplier theorem, Theorem 5.7.

We return to the rigorous arguments. For any $N \in \mathbb{I}$, we have

$$N \geq \max(2^{2^{\max(l,s_1,s_2)/C_0}}, C_3), \tag{7.1}$$

which implies, in particular, that

$$N \geq 2^{10du}. \tag{7.2}$$

In contrast, by Lemma 5.2(ii), $(\mathbb{Q}/\mathbb{Z})_{\le l}$ is the union of dual cyclic groups $\frac{1}{q}\mathbb{Z}/\mathbb{Z}$ with

$$q \le 2^{u/10}. \tag{7.3}$$

Thus N is going to be far larger than any single denominator q arising in the major arcs. If one wishes to contain $(\mathbb{Q}/\mathbb{Z})_{\le l}$ in a single dual cyclic group $\frac{1}{Q}\mathbb{Z}/\mathbb{Z}$, Lemma 5.2(ii) permits one to do this with

$$Q = Q_{\le l} \le 2^{2^{u/10}}. \tag{7.4}$$

Thus N may or may not be significantly larger than this Q. We will later separate N into large and small scales in order to exploit this containment in the large scale case.

From (7.2) we also have

$$-\operatorname{Log} N + l_{(N)} < -10u.$$

From (5.8) we see that the pair $(l, -u)$ has good major arcs. This lets us factor the expressions F_N, G_N using the symbol calculus (5.6). Indeed, if we set

$$F := \Pi_{l_1,\le -u} f \quad \text{and} \quad G := \Pi_{l_2,\le -u} g,$$

then from (5.15), (5.27), and (5.28) we have the identities

$$F_N = \mathrm{T}^{l_1}_{\varphi_N} F; \quad G_N = \mathrm{T}^{l_2}_{\tilde\varphi_N} G,$$

where $\varphi_N, \tilde\varphi_N \in \boldsymbol{S}(\mathbb{R})$ are the bump functions

$$\varphi_N(\xi) := \begin{cases} \eta(2^{\operatorname{Log} N - s_1}\xi) - \eta(2^{\operatorname{Log} N - s_1 + 1}\xi) & s_1 > -u, \\ \eta(2^{\operatorname{Log} N + u}\xi) & s_1 = -u \end{cases}$$

and

$$\tilde\varphi_N(\xi) := \begin{cases} \eta(2^{d(\operatorname{Log} N - s_2)}\xi) - \eta(2^{d(\operatorname{Log} N - s_2 + 1)}\xi) & s_2 > -u, \\ \eta(2^{d(\operatorname{Log} N + u)}\xi) & s_2 = -u. \end{cases} \tag{7.5}$$

From Lemma 5.17 we have

$$\|F\|_{\ell^p(\mathbb{Z})} \lesssim \langle l\rangle \|f\|_{\ell^p(\mathbb{Z})}; \quad \|G\|_{\ell^{p'}(\mathbb{Z})} \lesssim \langle l\rangle \|g\|_{\ell^{p'}(\mathbb{Z})}, \tag{7.6}$$

hence we may replace f, g by F, G respectively in the definition of acceptability. It will now suffice to show that the tuple

$$(\tilde A_N(\mathrm{T}^{l_1}_{\varphi_N} F, \mathrm{T}^{l_2}_{\tilde\varphi_N} G))_{N\in\mathbb{I}} \tag{7.7}$$

is acceptable.

The dependence on N has not yet materially improved, as the quantity $\tilde A_N(\mathrm{T}^{l_1}_{\varphi_N} F, \mathrm{T}^{l_2}_{\tilde\varphi_N} G)$ still depends on N in three different ways. However, we can

clarify the dependence on N by (adelic) Fourier analysis. From Example 4.10 and (4.9), we see that

$$\tilde{A}_N(\mathrm{T}^{l_1}_{\varphi_N} F, \mathrm{T}^{l_2}_{\tilde{\varphi}_N} G) = \mathrm{B}_{\mathcal{P}^{\otimes 2} m_N^{l_1,l_2}}(F,G),$$

where the symbol $m_N^{l_1,l_2}\colon (\mathbb{R}\times\mathbb{Q}/\mathbb{Z})^2 \to \mathbb{C}$ is defined by the formula

$$\begin{aligned} &m_N^{l_1,l_2}((\xi_1,\alpha_1),(\xi_2,\alpha_2)) \\ &:= \mathbb{1}_{\mathrm{h}(\alpha_1)=2^{l_1}} \mathbb{1}_{\mathrm{h}(\alpha_2)=2^{l_2}} \varphi_N(\xi_1)\tilde{\varphi}_N(\xi_2) \mathbb{E}_{n\in[N]} e((\alpha_1+\xi_1)n + (\alpha_2+\xi_2)P(n)) \mathbb{1}_{n>N/2}. \end{aligned}$$

From (7.2) and (7.3), we see that N is large compared to the naive heights of α_1,α_2, while $\xi_1',\xi_2' = O(2^{-u})$ are small on the support of $m_N^{l_1,l_2}$. This suggests that in the regimes of interest the symbol

$$\mathbb{E}_{n\in[N]} e((\alpha_1+\xi_1)n + (\alpha_2+\xi_2)P(n)) \mathbb{1}_{n>N/2}$$

has an approximate factorization

$$m_{\hat{\mathbb{Z}}}(\alpha_1,\alpha_2)\tilde{m}_{N,\mathbb{R}}(\xi_1,\xi_2), \tag{7.8}$$

where $m_{\hat{\mathbb{Z}}}\colon (\mathbb{Q}/\mathbb{Z})^2 \to \mathbb{C}$ is the normalized exponential sum

$$m_{\hat{\mathbb{Z}}}(\alpha_1,\alpha_2) := \int_{\hat{\mathbb{Z}}} e(\alpha_1 x + \alpha_2 P(x))\, d\mu_{\hat{\mathbb{Z}}}(x),$$

where $\mu_{\hat{\mathbb{Z}}}$ is the probability Haar measure on the profinite integers $\hat{\mathbb{Z}}$, or equivalently

$$m_{\hat{\mathbb{Z}}}\left(\frac{a_1}{q} \bmod 1, \frac{a_2}{q} \bmod 1\right) = \mathbb{E}_{n\in\mathbb{Z}/q\mathbb{Z}} e\left(\frac{a_1 n + a_2 P(n)}{q}\right)$$

for any $q\in\mathbb{Z}_+$ and $a_1,a_2\in\mathbb{Z}$, and $\tilde{m}_{N,\mathbb{R}}\colon \mathbb{R}^2\to\mathbb{C}$ is the oscillatory integral

$$\tilde{m}_{N,\mathbb{R}}(\xi_1,\xi_2) := \frac{1}{N}\int_{N/2}^{N} e(\xi_1 t + \xi_2 P(t))\, dt = \int_{1/2}^{1} e(\xi_1 N t + \xi_2 P(Nt))\, dt. \tag{7.9}$$

Note how the use of the upper averaging operators $\tilde{A}_N$ instead of A_N allows us to keep t bounded away from zero, which will be technically convenient later in the argument when we integrate by parts in t (as we now avoid the stationary points of P). The approximation (7.8) can be compared with (1.13).

The heuristic (7.8) then suggests the adelic bilinear symbol $m_N^{l_1,l_2} \in \boldsymbol{S}((\mathbb{R}\times\mathbb{Q}/\mathbb{Z})^2)$ approximately factors into the tensor product of a continuous bilinear symbol

$$(\varphi_N\otimes\tilde{\varphi}_N)\tilde{m}_{N,\mathbb{R}} \in \boldsymbol{S}(\mathbb{R}^2)$$

and the arithmetic bilinear symbol

$$m_{l_1,l_2,\hat{\mathbb{Z}}} := (\mathbb{1}_{(\mathbb{Q}/\mathbb{Z})_{l_1}} \otimes \mathbb{1}_{(\mathbb{Q}/\mathbb{Z})_{l_2}}) m_{\hat{\mathbb{Z}}} \in \boldsymbol{S}((\mathbb{Q}/\mathbb{Z})^2).$$

At the level of bilinear Fourier multipliers, this factorization suggests the approximation

$$\tilde{A}_N(\mathrm{T}^{l_1}_{\varphi_N} F, \mathrm{T}^{l_2}_{\tilde{\varphi}_N} G) \approx \mathrm{B}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}_{(\varphi_N \otimes \tilde{\varphi}_N)\tilde{m}_{N,\mathbb{R}}}(F,G),$$

where we introduce the twisted bilinear Fourier multiplier operators

$$\mathrm{B}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}_m := \mathrm{B}_{\mathcal{P}^{\otimes 2}(m \otimes m_{l_1,l_2,\hat{\mathbb{Z}}})} \tag{7.10}$$

for any $m \in \boldsymbol{S}(\mathbb{R}^2)$. More explicitly, one has

$$\mathrm{B}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}_m(f,g)(x) = \sum_{\alpha_1 \in (\mathbb{Q}/\mathbb{Z})_{l_1}, \alpha_2 \in (\mathbb{Q}/\mathbb{Z})_{l_2}} m_{\hat{\mathbb{Z}}}(\alpha_1,\alpha_2)$$
$$\times \int_{\mathbb{R}^2} m(\xi_1,\xi_2) \mathcal{F}_{\mathbb{Z}} f(\alpha_1+\xi_1) \mathcal{F}_{\mathbb{Z}} g(\alpha_2+\xi_2) e(-x(\alpha_1+\alpha_2+\xi_1+\xi_2))\, d\xi_1 d\xi_2.$$

Remark 7.11. Another way to think about the approximation (7.8) is that it is approximating the discrete averaging operator $\tilde{A}_N \colon \boldsymbol{S}(\mathbb{Z}) \times \boldsymbol{S}(\mathbb{Z}) \to \boldsymbol{S}(\mathbb{Z})$ by the adelic averaging operator $\tilde{A}_{N,\mathbb{A}_\mathbb{Z}} \colon \boldsymbol{S}(\mathbb{A}_\mathbb{Z}) \times \boldsymbol{S}(\mathbb{A}_\mathbb{Z}) \to \boldsymbol{S}(\mathbb{A}_\mathbb{Z})$ defined by

$$\tilde{A}_{N,\mathbb{A}_\mathbb{Z}}(f,g)(x) := \frac{1}{N} \int_{[N/2,N] \times \hat{\mathbb{Z}}} f(x-y) g(x-P(y))\, d\mu_{\mathbb{A}_\mathbb{Z}}(y), \tag{7.12}$$

which is in turn the tensor product of the continuous averaging operator $\tilde{A}_{N,\mathbb{R}} \colon \boldsymbol{S}(\mathbb{R}) \times \boldsymbol{S}(\mathbb{R}) \to \boldsymbol{S}(\mathbb{R})$ defined by

$$\tilde{A}_{N,\mathbb{R}}(f,g)(x) := \frac{1}{N} \int_{N/2}^{N} f(x-t) g(x-P(t))\, dt,$$

and the arithmetic averaging operator $A_{\hat{\mathbb{Z}}} \colon \boldsymbol{S}(\hat{\mathbb{Z}}) \times \boldsymbol{S}(\hat{\mathbb{Z}}) \to \boldsymbol{S}(\hat{\mathbb{Z}})$ defined in Example 4.10. As we shall see, this approximation is particularly accurate in the large-scale regime when N is large compared to the quantity $Q_{\leq l}$; see (A.4). In fact, the main estimate (3.10) on the integers $\mathbb{Z}$ has a natural analogue on the adelic integers $\mathbb{A}_\mathbb{Z}$ that can be proven by the same methods (with several simplifications). Our proof of the integer estimate was discovered by first working with the adelic operator (or more precisely, a projection of this operator to $\mathbb{R} \times \mathbb{Z}/Q\mathbb{Z}$) as a model case. This suggests that a natural route to prove other harmonic analysis estimates on the integers $\mathbb{Z}$ is to first study the analogous estimates on $\mathbb{A}_\mathbb{Z}$ or $\mathbb{R} \times \mathbb{Z}/Q\mathbb{Z}$ as model cases, in order to exploit the tensor product structure.

We now make the above heuristic precise. For future applications, we make the approximation slightly more general than what is needed in the current step.

PROPOSITION 7.13 (Major arc approximation of $\tilde{A}_N$). *For any $N \geq 1$ and $s \in \mathbb{N}$ with $-\operatorname{Log} N + s \leq -u$, we have*

$$\begin{aligned} &\Big\| \tilde{A}_N \left(\Pi_{l_1, \leq -\operatorname{Log} N + s} \tilde{F}, \Pi_{l_2, \leq -d \operatorname{Log} N + ds} \tilde{G} \right) \\ &\quad - \mathrm{B}^{l_1, l_2, m_{\hat{\mathbb{Z}}}}_{(\eta_{\leq -\operatorname{Log} N + s} \otimes \eta_{\leq -d \operatorname{Log} N + ds}) \tilde{m}_{N,\mathbb{R}}} (\tilde{F}, \tilde{G}) \Big\|_{\ell^p(\mathbb{Z})} \\ &\quad \lesssim_{C_3} 2^{O(\max(2^{\rho l}, s))} N^{-1} \|\tilde{F}\|_{\ell^{p_1}(\mathbb{Z})} \|\tilde{G}\|_{\ell^{p_2}(\mathbb{Z})} \end{aligned} \tag{7.14}$$

for all $\tilde{F} \in \ell^{p_1}(\mathbb{Z}), \tilde{G} \in \ell^{p_2}(\mathbb{Z})$.

The key point here is the gain of N^{-1} on the right-hand side, which in practice will make any expression estimated using this proposition acceptable (with room to spare).

Proof. From the same sort of calculations used in the preceding heuristic discussion, we can expand the expression inside the norm of the left-hand side (7.14) as

$$\mathrm{B}_{\mathcal{P}^{\otimes 2} M}(\tilde{F}, \tilde{G}),$$

where the symbol $M \in \boldsymbol{S}((\mathbb{R} \times \mathbb{Q}/\mathbb{Z})^2)$ is defined by

$$\begin{aligned} &M((\alpha_1, \xi_1), (\alpha_2, \xi_2)) \\ &:= \mathbb{1}_{\mathrm{h}(\alpha_1) = 2^{l_1}} \mathbb{1}_{\mathrm{h}(\alpha_2) = 2^{l_2}} \eta_{\leq -\operatorname{Log} N + s}(\xi_1) \eta_{\leq -d \operatorname{Log} N + ds}(\xi_2) M_0((\alpha_1, \xi_1), (\alpha_2, \xi_2)) \end{aligned}$$

with

$$\begin{aligned} &M_0((\alpha_1, \xi_1), (\alpha_2, \xi_2)) \\ &:= \mathbb{E}_{n \in [N]} e(\alpha_1 n + \alpha_2 P(n)) e(\xi_1 n + \xi_2 P(n)) \mathbb{1}_{n > N/2} - m_{\hat{\mathbb{Z}}}(\alpha_1, \alpha_2) \tilde{m}_{N,\mathbb{R}}(\xi_1, \xi_2). \end{aligned}$$

Applying Lemma 4.20 with $r_1 := N^{-1}$ and $r_2 := N^{-d}$, Lemma 5.2(iii), and the triangle inequality, as well as the Leibniz rule, it now suffices to establish the bounds

$$\frac{\partial^{j_1}}{\partial \xi_1^{j_1}} \frac{\partial^{j_2}}{\partial \xi_2^{j_2}} M_0((\alpha_1, \xi_1), (\alpha_2, \xi_2)) \lesssim_{C_3} 2^{O(\max(2^{\rho l}, s))} N^{j_1 + dj_2 - 1}$$

for $0 \leq j_1, j_2 \leq 2$, $\alpha_1 \in (\mathbb{Q}/\mathbb{Z})_{l_1}$, $\alpha_2 \in (\mathbb{Q}/\mathbb{Z})_{l_2}$, and $\xi_1 = O(2^s/N)$, $\xi_2 = O(2^{ds}/N^d)$.

By Lemma 5.2(ii), the sequence $n \mapsto e(\alpha_1 n + \alpha_2 P(n))$ is periodic with some period $q = O_\rho(2^{O(2^{\rho l})})$. Splitting into residue classes modulo q, and evaluating the derivatives, it suffices by the triangle inequality to show that

$$\sum_{n \in [N] \backslash [N/2]} w(n) \mathbb{1}_{n = a \bmod q} - \frac{1}{q} \int_{N/2}^{N} w(t)\, dt \lesssim_{C_3} 2^{O(\max(2^{\rho l}, s))} N^{j_1 + dj_2}$$

for all $a \in [q]$, where

$$w(t) := e(\xi_1 t + \xi_2 P(t)) t^{j_1} P(t)^{j_2}.$$

It suffices to show that

$$w(n) - \frac{1}{q}\int_n^{n+q} w(t)\ dt \lesssim_{C_3} 2^{O(\max(2^{\rho l},s))} N^{j_1+dj_2-1}$$

for all $n \in [N]\backslash[N/2]$, since the claim then follows by summing over all $n \in [N]\backslash[N/2]$ with $n = a \bmod q$ and using the triangle inequality to estimate the remainder. By the fundamental theorem of calculus, it then suffices to establish the bound

$$\frac{d}{dt} w(t) \lesssim_{C_3} 2^{O(\max(2^{\rho l},s))} N^{j_1+dj_2-1}$$

for $t \sim N$; but this follows from the hypotheses $\xi_1 = O(2^s/N)$, $\xi_2 = O(2^{ds}/N^d)$, and direct calculation. □

Applying this proposition with $\tilde{F} := \mathrm{T}^{l_1}_{\varphi_N} F$, $\tilde{G} := \mathrm{T}^{l_2}_{\tilde{\varphi}_N} G$, and $s := \max(0, s_1, s_2) + 1$, and using the functional calculus and Lemma 5.17, we conclude that

$$\|\tilde{A}_N(\mathrm{T}^{l_1}_{\varphi_N} F, \mathrm{T}^{l_2}_{\tilde{\varphi}_N} G) - \mathrm{B}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}_{(\varphi_N\otimes\tilde{\varphi}_N)\tilde{m}_{N,\mathbb{R}}}(F,G)\|_{\ell^p(\mathbb{Z})} \lesssim_{C_3} 2^{O(\max(2^{\rho l},s_1,s_2))} N^{-1} \|F\|_{\ell^{p_1}(\mathbb{Z})} \|G\|_{\ell^{p_2}(\mathbb{Z})}.$$

From (7.1) we certainly have

$$2^{O(\max(2^{\rho l},s_1,s_2))} \sum_{N\in\mathbb{I}} N^{-1} \lesssim_{C_3} \langle \max(l,s_1,s_2)\rangle^{O(1)} 2^{O(\rho l) - c\max(l,s_1,s_2)} \mathbb{1}_{p_1=p_2=2},$$

and thus by (2.7) and (7.6), we see that the tuple

$$(\tilde{A}_N(\mathrm{T}^{l_1}_{\varphi_N} F, \mathrm{T}^{l_2}_{\tilde{\varphi}_N} G) - \mathrm{B}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}_{(\varphi_N\otimes\tilde{\varphi}_N)\tilde{m}_{N,\mathbb{R}}}(F,G))_{N\in\mathbb{I}}$$

is acceptable. Thus by the triangle inequality, the acceptability of (7.7) is equivalent to the acceptability of

$$(\mathrm{B}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}_{(\varphi_N\otimes\tilde{\varphi}_N)\tilde{m}_{N,\mathbb{R}}}(F,G))_{N\in\mathbb{I}}.$$

From (2.6), it suffices to prove the acceptability of the two subtuples

$$(\mathrm{B}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}_{(\varphi_N\otimes\tilde{\varphi}_N)\tilde{m}_{N,\mathbb{R}}}(F,G))_{N\in\mathbb{I}_\le}, \quad (\mathrm{B}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}_{(\varphi_N\otimes\tilde{\varphi}_N)\tilde{m}_{N,\mathbb{R}}}(F,G))_{N\in\mathbb{I}_>},$$

where

$$\mathbb{I}_\le := \{N \in \mathbb{I} : N \le 2^{2^u}\} \tag{7.15}$$

is the set of "small scales," and

$$\mathbb{I}_> := \{N \in \mathbb{I} : N > 2^{2^u}\}$$

is the set of "large scales." As we shall see, for the small scales one will be able to tolerate the (doubly) logarithmic losses arising from Rademacher–Menshov arguments, and for the large scales, one will be able to exploit (7.4) to replace the integers $\mathbb{Z}$ by the adelic integers $\mathbb{A}_{\mathbb{Z}}$.

At this stage the bilinear operator $\mathrm{B}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}_{(\varphi_N\otimes\tilde{\varphi}_N)\tilde{m}_{N,\mathbb{R}}}$ still has a symbol that depends on N, although at least the dependence is now confined to the continuous frequency variables and not the arithmetic ones. To simplify the dependence further, we observe from (4.17) that we have the functional calculus

$$\mathrm{B}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}_{(\varphi_1\otimes\varphi_2)m}(f,g) = \mathrm{B}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}_{m}(\mathrm{T}^{l_1}_{\varphi_1}f, \mathrm{T}^{l_2}_{\varphi_2}g)$$

whenever $\varphi_1,\varphi_2 \in \boldsymbol{S}(\mathbb{R}_{\leq -u})$ and $m \in \boldsymbol{S}(\mathbb{R}^2_{\leq -u})$. From this calculus and the definition (7.9) of $\tilde{m}_{N,\mathbb{R}}$, we can factor $\mathrm{B}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}_{(\varphi_N\otimes\tilde{\varphi}_N)\tilde{m}_{N,\mathbb{R}}}(F,G)$ as

$$\mathrm{B}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}_{(\varphi_N\otimes\tilde{\varphi}_N)\tilde{m}_{N,\mathbb{R}}}(F,G) = \int_{1/2}^{1} \mathrm{B}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}_{m_*}(\mathrm{T}^{l_1}_{\varphi_{N,t}}F, \mathrm{T}^{l_2}_{\tilde{\varphi}_{N,t}}G)\, dt, \tag{7.16}$$

where $\varphi_{N,t},\tilde{\varphi}_{N,t} \in \boldsymbol{S}(\mathbb{R})$ are modulated variants of $\varphi_N,\tilde{\varphi}_N$ defined by the formulae

$$\begin{aligned} \varphi_{N,t}(\xi) &\coloneqq \varphi_N(\xi)e(Nt\xi), \\ \tilde{\varphi}_{N,t}(\xi) &\coloneqq \tilde{\varphi}_N(\xi)e(P(Nt)\xi) \end{aligned} \tag{7.17}$$

and $m_* \in \boldsymbol{S}(\mathbb{R}^2)$ is the symbol

$$m_* \coloneqq \eta_{\leq -2u} \otimes \eta_{\leq -2du}.$$

The advantage of the formulation (7.16) is that the bilinear operator $\mathrm{B}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}_{m_*}$ is independent of N. This is particularly useful in the small-scale case $N \in \mathbb{I}_{\leq}$, as it will let us control variational norms of bilinear expressions in terms of linear quantities via a two-parameter version of the Rademacher–Menshov inequality.

In the large-scale case $N \in \mathbb{I}_{>}$, we can express (7.16) in another useful way. Introduce the adelic model functions $F_{\mathbb{A}} \in L^{p_1}(\mathbb{A}_{\mathbb{Z}})$, $G_{\mathbb{A}} \in L^{p_2}(\mathbb{A}_{\mathbb{Z}})$ by the formulae

$$F_{\mathbb{A}}(x,y) \coloneqq \sum_{\alpha_1 \in (\mathbb{Q}/\mathbb{Z})_{l_1}} \int_{\mathbb{R}} \eta_{\leq -2^{u-1}}(\xi_1)\mathcal{F}_{\mathbb{Z}}F(\alpha_1+\xi_1)e(-(\xi_1,\alpha_1)\cdot(x,y))\, d\xi_1 \tag{7.18}$$

and

$$G_{\mathbb{A}}(x,y) \coloneqq \sum_{\alpha_2 \in (\mathbb{Q}/\mathbb{Z})_{l_2}} \int_{\mathbb{R}} \eta_{\leq -2^{u-1}}(\xi_2)\mathcal{F}_{\mathbb{Z}}G(\alpha_2+\xi_2)e(-(\xi_2,\alpha_2)\cdot(x,y))\, d\xi_2 \tag{7.19}$$

for $x \in \mathbb{R}, y \in \hat{\mathbb{Z}}$, or equivalently on the Fourier side

$$\begin{aligned} \mathcal{F}_{\mathbb{A}_{\mathbb{Z}}}F_{\mathbb{A}}(\xi_1,\alpha_1) &= \mathbb{1}_{\mathrm{h}(\alpha_1)=2^{l_1}}\eta_{\leq -2^{u-1}}(\xi_1)\mathcal{F}_{\mathbb{Z}}F(\alpha_1+\xi_1), \\ \mathcal{F}_{\mathbb{A}_{\mathbb{Z}}}G_{\mathbb{A}}(\xi_2,\alpha_2) &= \mathbb{1}_{\mathrm{h}(\alpha_2)=2^{l_2}}\eta_{\leq -2^{u-1}}(\xi_2)\mathcal{F}_{\mathbb{Z}}G(\alpha_2+\xi_2) \end{aligned}$$

for $\xi_1, \xi_2 \in \mathbb{R}$ and $\alpha_1, \alpha_2 \in \mathbb{Q}/\mathbb{Z}$. (One can use Lemma 4.20 to verify that $F_{\mathbb{A}}$ does indeed lie in $L^p(\mathbb{A}_{\mathbb{Z}})$, and similarly for $G_{\mathbb{A}}$.) One can also interpret $F_{\mathbb{A}}, G_{\mathbb{A}}$ as the interpolated functions

$$F_{\mathbb{A}} = \mathcal{S}^{-1}_{\mathbb{R}_{\leq -2^{u-1}} \times (\mathbb{Q}/\mathbb{Z})_{l_1}} \Pi_{l_1, \leq -2^{u-1}} F, \quad G_{\mathbb{A}} = \mathcal{S}^{-1}_{\mathbb{R}_{\leq -2^{u-1}} \times (\mathbb{Q}/\mathbb{Z})_{l_2}} \Pi_{l_2, \leq -2^{u-1}} G.$$

In the large-scale case, $\eta_{\leq -2^{u-1}}$ equals 1 on the support of $\varphi_{N,t}$, $\tilde{\varphi}_{N,t}$, and m_* equals 1 on the support of $\eta_{\leq -2^{u-1}} \otimes \eta_{\leq -2^{u-1}}$. One can then describe various combinations of F, G as applications of the sampling operator $\mathcal{S}$ to various combinations of $F_{\mathbb{A}}, G_{\mathbb{A}}$. More precisely, one observes the identities

$$\Pi_{l_1, \leq -2^{u-1}} F = \mathcal{S} F_{\mathbb{A}}, \tag{7.20}$$

$$\Pi_{l_2, \leq -2^{u-1}} G = \mathcal{S} G_{\mathbb{A}}, \tag{7.21}$$

$$\mathrm{T}^{l_1}_{\varphi_{N,t}} F = \mathcal{S} \mathrm{T}_{\varphi_{N,t} \otimes 1} F_{\mathbb{A}},$$

$$\mathrm{T}^{l_2}_{\tilde{\varphi}_{N,t}} G = \mathcal{S} \mathrm{T}_{\tilde{\varphi}_{N,t} \otimes 1} G_{\mathbb{A}},$$

$$\mathrm{B}^{l_1, l_2, m_{\hat{\mathbb{Z}}}}_{m_*}(\mathrm{T}^{l_1}_{\varphi_{N,t}} F, \mathrm{T}^{l_2}_{\tilde{\varphi}_{N,t}} G) = \mathcal{S} \mathrm{B}_{1 \otimes m_{l_1, l_2, \hat{\mathbb{Z}}}}(\mathrm{T}_{\varphi_{N,t} \otimes 1} F_{\mathbb{A}}, \mathrm{T}_{\tilde{\varphi}_{N,t} \otimes 1} G_{\mathbb{A}})$$

so that (7.16) can now be written as

$$\mathcal{S} \int_{1/2}^1 \mathrm{B}_{(\varphi_{N,t} \otimes \tilde{\varphi}_{N,t}) \otimes m_{l_1, l_2, \hat{\mathbb{Z}}}}(F_{\mathbb{A}}, G_{\mathbb{A}})\, dt.$$

All functions on $\mathbb{A}_{\mathbb{Z}}$ here have Fourier support in the region

$$(\mathbb{R}_{\leq -2^{u-1}} \times (\mathbb{Q}/\mathbb{Z})_{l_1}) \times (\mathbb{R}_{\leq -2^{u-1}} \times (\mathbb{Q}/\mathbb{Z})_{l_2}),$$

which by Lemma 5.2(ii) is contained in

$$\Big(\mathbb{R}_{\leq -2^{u-1}} \times \Big(\frac{1}{Q_{\leq l}} \mathbb{Z}/\mathbb{Z}\Big)\Big) \times \Big(\mathbb{R}_{\leq -2^{u-1}} \times \Big(\frac{1}{Q_{\leq l}} \mathbb{Z}/\mathbb{Z}\Big)\Big).$$

In this large-scale regime, this is a regime in which Theorem 4.18 applies, thanks to (7.4). In particular, from Theorem 4.18 (using the normed vector space $\boldsymbol{V}^r$), we have

$$\|(\mathrm{B}^{l_1, l_2, m_{\hat{\mathbb{Z}}}}_{(\varphi_N \otimes \tilde{\varphi}_N) \tilde{m}_{N,\mathbb{R}}}(F, G))_{N \in \mathbb{I}_>}\|_{\ell^p(\mathbb{Z}; \boldsymbol{V}^r)}$$

$$\sim \Big\| \Big(\int_{1/2}^1 \mathrm{B}_{1 \otimes m_{\hat{\mathbb{Z}}}}(\mathrm{T}_{\varphi_{N,t} \otimes 1} F_{\mathbb{A}}, \mathrm{T}_{\tilde{\varphi}_{N,t} \otimes 1} G_{\mathbb{A}})\, dt \Big)_{N \in \mathbb{I}_>} \Big\|_{L^p(\mathbb{A}_{\mathbb{Z}}; \boldsymbol{V}^r))};$$

similarly, from (7.20), (7.21), (7.6), Theorem 4.18, and Lemma 5.17 one has

$$\|F_{\mathbb{A}}\|_{L^{p_1}(\mathbb{A}_{\mathbb{Z}})} \lesssim \langle l \rangle^{O(1)} \|f\|_{\ell^{p_1}(\mathbb{Z})}; \quad \|G_{\mathbb{A}}\|_{L^{p_2}(\mathbb{A}_{\mathbb{Z}})} \lesssim \langle l \rangle^{O(1)} \|g\|_{\ell^{p_2}(\mathbb{Z})}. \tag{7.22}$$

In view of the above discussion (and Proposition 5.33), Theorem 5.30 (and hence Theorem 1.17) now reduces to establishing the following estimates.

THEOREM 7.23 (Model operator estimates, I). *Suppose that at least one of $s_1 = -u$, $s_2 = -u$, or $p \neq 2$ holds. Then the small-scale model tuple*

$$\Big(\int_{1/2}^1 \mathrm{B}_{m_*}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}(\mathrm{T}^{l_1}_{\varphi_{N,t}} F, \mathrm{T}^{l_2}_{\tilde{\varphi}_{N,t}} G)\ dt \Big)_{N \in \mathbb{I}_{\leq}} \tag{7.24}$$

and the large-scale model tuple

$$\Big(\int_{1/2}^1 \mathrm{B}_{1 \otimes m_{\hat{\mathbb{Z}}}}(\mathrm{T}_{\varphi_{N,t} \otimes 1} F_{\mathbb{A}}, \mathrm{T}_{\tilde{\varphi}_{N,t} \otimes 1} G_{\mathbb{A}}) \Big)_{N \in \mathbb{I}_{>}} \tag{7.25}$$

are both acceptable.

It remains to establish Theorem 7.23. One difficulty in this theorem is the need to obtain some decay in s_1, s_2 when they are large. Our main tool for doing this will be the following integration by parts identity. For $j_1, j_2 = -1, 0, +1$ with $(s_1, j_1), (s_2, j_2) \neq (-u, -1)$, we define the modified bump functions

$$\varphi_{N,t,j_1}(\xi_1) := (2^{-s_1} N \xi_1)^{j_1} \varphi_{N,t}(\xi_1) = (2^{-s_1} N \xi_1)^{j_1} e(Nt\xi_1) \varphi_N(\xi_1)$$

and

$$\tilde{\varphi}_{N,t,j_2}(\xi_2) := (2^{-ds_2} N^d \xi_2)^{j_2} \tilde{\varphi}_{N,t}(\xi_2) = (2^{-ds_2} N^d \xi_2)^{j_2} e(P(Nt)\xi_2) \tilde{\varphi}_N(\xi_2).$$

Note that it is necessary to exclude the cases $(s_1, j_1), (s_2, j_2) = (-u, -1)$ to prevent these functions from developing a singularity at the frequency origin.

LEMMA 7.26 (Integration by parts identity).

(i) *If $s_1 > -u$, then we have*

$$\begin{aligned} &\int_{1/2}^1 \varphi_{N,t} \otimes \tilde{\varphi}_{N,t}\ dt \\ &\quad = \frac{2^{-s_1}}{2\pi i} \varphi_{N,t,-1} \otimes \tilde{\varphi}_{N,t} \Big|_{t=1/2}^{t=1} - 2^{ds_2 - s_1} \int_{1/2}^1 \varphi_{N,t,-1} \otimes \tilde{\varphi}_{N,t,1} \frac{P'(Nt)}{N^{d-1}}\ dt. \end{aligned}$$

(ii) *If $s_2 > -u$, then we have*

$$\begin{aligned} \int_{1/2}^1 \varphi_{N,t} \otimes \tilde{\varphi}_{N,t}\ dt &= \frac{2^{-ds_2}}{2\pi i} \varphi_{N,t} \otimes \tilde{\varphi}_{N,t,-1} \frac{N^{d-1}}{P'(Nt)} \Big|_{t=1/2}^{t=1} \\ &\quad - 2^{s_1 - ds_2} \int_{1/2}^1 \varphi_{N,t,1} \otimes \tilde{\varphi}_{N,t,-1} \frac{N^{d-1}}{P'(Nt)}\ dt \\ &\quad + \frac{2^{-ds_2}}{2\pi i} \int_{1/2}^1 \varphi_{N,t} \otimes \tilde{\varphi}_{N,t,-1} \frac{N^d P''(Nt)}{P'(Nt)^2}\ dt. \end{aligned}$$

Note that the quantity $P'(Nt)$ that appears in some of the denominators here is non-vanishing thanks to the lower bounds $N \geq C_3$ and $t \geq 1/2$; indeed

the tuples

$$\left(\frac{P'(Nt)}{N^{d-1}}\right)_{N\in\mathbb{I}}, \left(\frac{N^{d-1}}{P'(Nt)}\right)_{N\in\mathbb{I}}, \left(\frac{N^d P''(Nt)}{P'(Nt)^2}\right)_{N\in\mathbb{I}} \tag{7.27}$$

can all be easily verified to have a $\boldsymbol{V}^r$ norm of $O(1)$ for all $1/2 \le t \le 1$. This is the main reason why we work with $\tilde{A}_N$ instead of A_N in most of this paper.

Proof. To prove (i) it suffices to show that

$$\int_{1/2}^1 e(\xi_1 Nt + \xi_2 P(Nt))\, dt = \frac{e(\xi_1 Nt + \xi_2 P(Nt))}{2\pi i N \xi_1}\Big|_{t=1/2}^{t=1} - \int_{1/2}^1 e(\xi_1 Nt + \xi_2 P(Nt)) \frac{P'(Nt)\xi_2}{\xi_1}\, dt$$

whenever $\xi_1 \neq 0$ and $N \ge C_3$, but this follows by writing

$$e(\xi_1 Nt) = \frac{1}{2\pi i N \xi_1}\frac{d}{dt} e(\xi_1 Nt)$$

and integrating by parts. Similarly, to prove (ii) it suffices to show that

$$\int_{1/2}^1 e(\xi_1 Nt + \xi_2 P(Nt))\, dt = \frac{e(\xi_1 Nt + \xi_2 P(Nt))}{2\pi i N \xi_2 P'(Nt)}\Big|_{t=1/2}^{t=1} - \int_{1/2}^1 e(\xi_1 Nt + \xi_2 P(Nt)) \frac{\xi_1}{P'(Nt)\xi_2}\, dt + \frac{1}{2\pi i}\int_{1/2}^1 e(\xi_1 Nt + \xi_2 P(Nt)) \frac{P''(Nt)}{\xi_2 P'(Nt)^2}\, dt$$

whenever $\xi_2 \neq 0$ and $N \ge C_3$, but this follows by writing $e(\xi_2 P(Nt)) = \frac{1}{2\pi i N \xi_2 P'(Nt)}\frac{d}{dt} e(\xi_2 P(Nt))$ and integrating by parts. □

We will now show how Theorem 7.23 is a consequence of Lemma 7.26 and the following variant, which works with a fixed choice of t but does not require any decay in the s_1, s_2 parameters.

THEOREM 7.28 (Model operator estimates, II). *Let $j_1, j_2 \in \{-1, 0, +1\}$ be such that*

$$(s_1, j_1), (s_2, j_2) \neq (-u, -1). \tag{7.29}$$

Then for every $1/2 \le t \le 1$, one has the small-scale model estimate

$$\Big\| \Big(\mathrm{B}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}_{m_*}(\mathrm{T}^{l_1}_{\varphi_{N,t,j_1}} F, \mathrm{T}^{l_2}_{\tilde{\varphi}_{N,t,j_2}} G) \Big)_{N\in\mathbb{I}_\le} \Big\|_{\ell^p(\mathbb{Z};\boldsymbol{V}^r)} \lesssim_{C_3} \langle \max(l, s_1, s_2)\rangle^{O(1)} 2^{O(\rho l) - cl\mathbb{1}_{p_1=p_2=2}} \|F\|_{\ell^{p_1}(\mathbb{Z})} \|G\|_{\ell^{p_2}(\mathbb{Z})} \tag{7.30}$$

and the large-scale model estimate

$$\begin{aligned}(7.31)\qquad &\Big\|\big(\mathrm{B}_{1\otimes m_{\hat{\mathbb{Z}}}}(\mathrm{T}_{\varphi_{N,t,j_1}\otimes 1}F_{\mathbb{A}},\mathrm{T}_{\tilde{\varphi}_{N,t,j_2}\otimes 1}G_{\mathbb{A}})\big)_{N\in\mathbb{I}_>}\Big\|_{L^p(\mathbb{A}_{\mathbb{Z}};\boldsymbol{V}^r)}\\ &\lesssim_{C_3}\langle\max(l,s_1,s_2)\rangle^{O(1)}2^{O(\rho l)-cl\mathbb{1}_{p_1=p_2=2}}\|F_{\mathbb{A}}\|_{L^{p_1}(\mathbb{A}_{\mathbb{Z}})}\|G_{\mathbb{A}}\|_{L^{p_2}(\mathbb{A}_{\mathbb{Z}})}.\end{aligned}$$

We assume Theorem 7.28 for now and show how it implies Theorem 7.23. We give the argument for the large-scale tuple (7.25), as the treatment of the small-scale tuple (7.24) is completely analogous. From Theorem 7.28 (with $j_1=j_2=0$), (7.6), (7.22) and Minkowski's integral inequality, we already obtain the acceptability bound for (7.25) but with the factor $2^{-c\max(l,s_1,s_2)\mathbb{1}_{p_1=p_2=2}}$ replaced by $2^{-cl\mathbb{1}_{p_1=p_2=2}}$. This gives the claim unless $p_1=p_2=2$ and $\max(s_1,s_2)>l$ so, in particular, $p=1$. Since the high-high case $s_1,s_2>-u$, $p_1=p_2=2$ has already been excluded, this only leaves us with the high-low case $s_1>l$, $s_2=-u$, $p_1=p_2=2$ and the low-high case $s_2>l$, $s_1=-u$, $p_1=p_2=2$. In the low-high case, one applies Lemma 7.26(ii), (7.27), (2.8), and Minkowski's integral inequality to bound the left-hand side of (7.25) (where the integrand can be viewed as a linear functional applied to $\varphi_{N,t}\otimes\tilde{\varphi}_{N,t}$) by

$$\lesssim 2^{-ds_2}\sup_{j_1,j_2=0,\pm1}\sup_{1/2\le t\le 1}\|(\mathrm{B}_{1\otimes m_{\hat{\mathbb{Z}}}}(\mathrm{T}_{\varphi_{N,t,j_1}\otimes 1}F_{\mathbb{A}},\mathrm{T}_{\tilde{\varphi}_{N,t,j_2}\otimes 1}G_{\mathbb{A}}))_{N\in\mathbb{I}_>}\|_{L^1(\mathbb{A}_{\mathbb{Z}};\boldsymbol{V}^r)}.$$

The acceptability of (7.25) in this case now follows from Theorem 7.28, (7.6) and (7.22) (noting that the hypothesis (7.29) is verified). In the high-low case one argues similarly using Lemma 7.26(i) instead of Lemma 7.26(ii).

It remains to establish Theorem 7.28. This will be the purpose of the next three sections of this paper.

8. The small-scale estimate: applying the Rademacher–Menshov inequality

In this section we establish (7.30). A key tool in the small-scale case will be the following two-dimensional version of the Rademacher–Menshov inequality.

LEMMA 8.1 (Two-dimensional Rademacher–Menshov). *Let $K\in\mathbb{Z}_+$, and for any $k_1,k_2\in[K]$, let a_{k_1,k_2} be a complex number, with the convention that $a_{k_1,k_2}=0$ if $k_1=0$ or $k_2=0$. Then for any $1<r<\infty$, one has*

$$\|(a_{k,k})_{k\in[K]}\|_{\boldsymbol{V}^r}\lesssim_r\sum_{M_1,M_2\in 2^{\mathbb{N}}\cap[K]}\big\|(\Delta a_{M_1j_1,M_2j_2})_{(j_1,j_2)\in[K/M_1]\times[K/M_2]}\big\|_{\ell^r},$$

where

$$\Delta a_{M_1j_1,M_2j_2}:=a_{M_1j_1,M_2j_2}-a_{M_1(j_1-1),M_2j_2}-a_{M_1j_1,M_2(j_2-1)}+a_{M_1(j_1-1),M_2(j_2-1)}.$$

The one-dimensional analogue of this inequality is well known; see, e.g., [67, Lemma 2.5, pp. 534].

Proof. By definition (2.5) of the $\boldsymbol{V}^r$ norm, one has

$$\|(a_{k,k})_{k\in[K]}\|_{\boldsymbol{V}^r} \lesssim_r \|(a_{k_j,k_j} - a_{k_{j-1},k_{j-1}})_{j\in[J]}\|_{\ell^r}$$

for some sequence $1 \le k_1 < \cdots < k_J \le K$, with the convention $k_0 = 0$.

Let μ be the discrete complex measure on $[K]^2$ with masses

$$\mu(\{(l_1,l_2)\}) := a_{l_1,l_2} - a_{l_1-1,l_2} - a_{l_1,l_2-1} + a_{l_1-1,l_2-1}.$$

By the telescoping series we may write

$$a_{k_j,k_j} - a_{k_{j-1},k_{j-1}} = \mu([k_j]^2\backslash[k_{j-1}]^2).$$

Observe that the L-shaped region $[k_j]^2\backslash[k_{j-1}]^2$ can be partitioned into the union of two rectangles:

$$[k_j]^2\backslash[k_{j-1}]^2 = [k_j] \times ([k_j]\backslash[k_{j-1}]) \uplus ([k_j]\backslash[k_{j-1}]) \times [k_{j-1}].$$

We partition these rectangles further into dyadic subrectangles as follows. For each $M \in 2^{\mathbb{N}} \cap [K]$, let $\mathcal{I}_M$ be the collection of all discrete dyadic intervals I in $[K]$ of length M, thus $I = [M] + (j-1)M = \{jM - M + 1, \dots, M\}$ for some $j \in [K/M]$. Every interval J in $[K]$ can then be written as the union of disjoint dyadic intervals $I \in \bigcup_{M\in 2^{\mathbb{N}}\cap[K]} \mathcal{I}_M$, in such a manner that at most two intervals are used from each collection $\mathcal{I}_M$. Indeed, one can take the I to be the maximal dyadic intervals contained in J: for each scale M, the intervals in $\mathcal{I}_M$ that lie in J are consecutive, and all but the two extreme intervals in this sequence will fail to be maximal. Taking Cartesian products, we conclude that the region $[k_j]^2\backslash[k_{j-1}]^2$ can be written as the union of dyadic rectangles $I_1 \times I_2$ with $I_1 \in \mathcal{I}_{M_1}, I_2 \in \mathcal{I}_{M_2}$ for some $M_1, M_2 \in 2^{\mathbb{N}} \cap [K]$, in such a way that each pair (M_1, M_2) is associated to $O(1)$ rectangles $I_1 \times I_2$. From the triangle inequality, we thus have

$$\mu([k_j]^2\backslash[k_{j-1}]^2) \lesssim \sum_{M_1,M_2\in 2^{\mathbb{N}}\cap[K]} \sup_{I_1\in\mathcal{I}_{M_1}, I_2\in\mathcal{I}_{M_2}: I_1\times I_2 \subset [k_j]^2\backslash[k_{j-1}]^2} |\mu(I_1 \times I_2)|$$

and hence on taking ℓ^r norms

$$\|(a_{k,k})_{k\in[K]}\|_{\boldsymbol{V}^r} \lesssim_r \sum_{M_1,M_2\in 2^{\mathbb{N}}\cap[K]} \Big\| \Big(\sup_{I_1\in\mathcal{I}_{M_1}, I_2\in\mathcal{I}_{M_2}: I_1\times I_2 \subset [k_j]^2\backslash[k_{j-1}]^2} |\mu(I_1 \times I_2)| \Big)_{j\in[J]} \Big\|_{\ell^r};$$

since the rectangles $I_1 \times I_2$ associated to a given region $[k_j]^2\backslash[k_{j-1}]^2$ are disjoint, we conclude that

$$\|(a_{k,k})_{k\in[K]}\|_{\boldsymbol{V}^r} \lesssim_r \sum_{M_1,M_2\in 2^{\mathbb{N}}\cap[K]} \Big\| (\mu(I_1 \times I_2))_{I_1\in\mathcal{I}_{M_1}, I_2\in\mathcal{I}_{M_2}} \Big\|_{\ell^r}.$$

If $I_1 = [M_1] + (j_1 - 1)M_1$ and $I_2 = [M_2] + (j_2 - 1)M_2$, then

$$\mu(I_1 \times I_2) = a_{M_1 j_1, M_2 j_2} - a_{M_1(j_1-1), M_2 j_2} - a_{M_1 j_1, M_2(j_2-1)} + a_{M_1(j_1-1), M_2(j_2-1)}$$

and the claim follows. □

We can combine this with Khintchine's inequality to conclude

COROLLARY 8.2 (Rademacher–Menshov for bilinear forms). *Let $K \in \mathbb{Z}_+$, and for any $k \in [K]$, let $f_k \in V, g_k \in W$ be elements of some vector spaces V, W. Let $0 < q < \infty$, and let $B : V \times W \to L^q(X)$ be a bilinear map for some measure space X. Then*

(8.3)

$$\|(B(f_k, g_k))_{k\in[K]}\|_{L^q(X;V^2)} \lesssim_q \langle \log K \rangle^{\max(2, \frac{2}{q})} \times \sup_{\epsilon_1, \epsilon'_1, \dots, \epsilon_K, \epsilon'_K \in \{-1,+1\}} \Big\| B\Big(\sum_{k\in[K]} \epsilon_k (f_k - f_{k-1}), \sum_{k\in[K]} \epsilon'_k (g_k - g_{k-1}) \Big) \Big\|_{L^q(X)}$$

with the conventions $f_0 = g_0 = 0$.

In our applications, the set $[K]$ will index a lacunary set of scales, so the $\log K$ type losses are in fact doubly logarithmic in the scale parameters. This will allow us to profitably use this corollary for scales as large as 2^{2^u}. Note in this corollary that the bilinear operator B is not permitted to depend on k, but fortunately the Fourier-analytic manipulations of the preceding section have achieved such an independence of k for the bilinear operator appearing in (7.30).

Proof. We may normalize

(8.4)

$$\sup_{\epsilon_1, \epsilon'_1, \dots, \epsilon_K, \epsilon'_K \in \{-1,+1\}} \Big\| B\Big(\sum_{k\in[K]} \epsilon_k (f_k - f_{k-1}), \sum_{k\in[K]} \epsilon'_k (g_k - g_{k-1}) \Big) \Big\|_{L^q(X)} = 1.$$

For each $x \in X$, we apply Lemma 8.1 with $a_{k_1,k_2} = B(f_{k_1}, g_{k_2})(x)$ and $r = 2$ to bound the left-hand side of (8.3) by

$$\lesssim \Big\| \sum_{M_1, M_2 \in 2^{\mathbb{N}} \cap [K]} \big\| (B(\tilde f_{M_1 j_1}, \tilde g_{M_2 j_2}))_{(j_1,j_2)\in[K/M_1]\times[K/M_2]} \big\|_{\ell^2} \Big\|_{L^q(X)},$$

where $\tilde f_{M_1 j_1} := f_{M_1 j_1} - f_{M_1(j_1-1)}$ and $\tilde g_{M_2 j_2} := g_{M_2 j_2} - g_{M_2(j_2-1)}$. The last norm by the triangle or quasi–triangle inequality (3.7) is bounded by

$$\lesssim_q \langle K \rangle^{\max(2, \frac{2}{q})} \sup_{M_1, M_2 \in 2^{\mathbb{N}} \cap [K]} \Big\| (B(\tilde f_{M_1 j_1}, \tilde g_{M_2 j_2}))_{(j_1,j_2)\in[K/M_1]\times[K/M_2]} \Big\|_{L^q(X;\ell^2)}.$$

Thus it suffices to show for each $M_1, M_2 \in 2^{\mathbb{N}} \cap [K]$ that

$$\Big\| (B(\tilde f_{M_1 j_1}, \tilde g_{M_2 j_2}))_{(j_1,j_2)\in[K/M_1]\times[K/M_2]} \Big\|^q_{L^q(X;\ell^2)} \lesssim_q 1.$$

But by two applications of Khintchine's inequality, one can bound the left-hand side by the expected value of

$$\Big\| \sum_{j_1 \in [K/M_1]} \sum_{j_2 \in [K/M_2]} \epsilon_{j_1} \epsilon'_{j_2} B(\tilde{f}_{M_1 j_1}, \tilde{g}_{M_2 j_2}) \Big\|_{L^q(X)}^q,$$

where $\epsilon_{j_1}, \epsilon'_{j_2}$ are independent random Bernoulli signs. But every instance of this random expression can be factored (after relabeling the signs) in the form of one of the norms in (8.4), raised to the power q, and the claim follows. □

We now apply this estimate to (7.30). We enumerate the elements of $\mathbb{I}_{\leq}$ in order as $N_1 < \cdots < N_K$; we may assume that $K \geq 1$ since otherwise there is nothing to prove. From (7.15) we have $K = O(2^u)$. Thus by Lemma 8.2 we may bound the left-hand side of (7.30) by

$$u^{O(1)} \| \mathrm{B}_{m_*}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}(\mathrm{T}_{\varphi_*}^{l_1} F, \mathrm{T}_{\tilde{\varphi}_*}^{l_2} G) \|_{\ell^p(\mathbb{Z})}$$

for some cutoffs $\varphi_*, \tilde{\varphi}_*$ of the form

$$\begin{aligned} \varphi_* &= \sum_{k \in [K]} \epsilon_k (\varphi_{N_k,t,j_1} - \varphi_{N_{k-1},t,j_1}), \\ \tilde{\varphi}_* &= \sum_{k \in [K]} \tilde{\epsilon}_k (\tilde{\varphi}_{N_k,t,j_2} - \tilde{\varphi}_{N_{k-1},t,j_2}) \end{aligned} \tag{8.5}$$

for some signs $\epsilon_k, \tilde{\epsilon}_k \in \{-1, +1\}$, where we adopt the convention $\varphi_{N_0,t,j_1} = \tilde{\varphi}_{N_0,t,j_2} = 0$. Note from (5.26) that $u^{O(1)} \lesssim_{C_3} 2^{O(\rho l)}$, so the loss of $u^{O(1)}$ will be acceptable for us. It now suffices to show that

$$\begin{aligned} &\| \mathrm{B}_{m_*}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}(\mathrm{T}_{\varphi_*}^{l_1} F, \mathrm{T}_{\tilde{\varphi}_*}^{l_2} G) \|_{\ell^p(\mathbb{Z})} \\ &\quad \lesssim_{C_3} \langle \max(l, s_1, s_2) \rangle^{O(1)} 2^{-cl \mathbb{1}_{p_1 = p_2 = 2}} \|F\|_{\ell^{p_1}(\mathbb{Z})} \|G\|_{\ell^{p_2}(\mathbb{Z})}. \end{aligned}$$

We now use

Lemma 8.6 (Single-scale estimate). *If $\tilde{F} \in \ell^{p_1}(\mathbb{Z}), \tilde{G} \in \ell^{p_2}(\mathbb{Z})$ have Fourier support on $\mathcal{M}_{l_1,\leq -3u}$ and $\mathcal{M}_{l_2,\leq -3du}$ respectively, then*

$$\| \mathrm{B}_{m_*}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}(\tilde{F}, \tilde{G}) \|_{\ell^p(\mathbb{Z})} \lesssim_{C_3} 2^{-cl \mathbb{1}_{p_1 = p_2 = 2}} \|\tilde{F}\|_{\ell^{p_1}(\mathbb{Z})} \|\tilde{G}\|_{\ell^{p_2}(\mathbb{Z})}.$$

Proof. The strategy is to apply Proposition 7.13 in reverse, so that Theorem 5.12 may be applied. We may normalize $\|\tilde{F}\|_{\ell^{p_1}(\mathbb{Z})} = \|\tilde{G}\|_{\ell^{p_2}(\mathbb{Z})} = 1$. From Proposition 7.13 with $N = 2^u$ and $s = 0$, we see that

$$\| \tilde{A}_{2^u}(\tilde{F}, \tilde{G}) - \mathrm{B}_{\tilde{m}_{2^u,\mathbb{R}} m_*}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}(\tilde{F}, \tilde{G}) \|_{\ell^p(\mathbb{Z})} \lesssim_{C_3} 2^{O(2^{\rho l}) - u} \lesssim_{C_3} 2^{-cl \mathbb{1}_{p_1 = p_2 = 2}},$$

noting that on the Fourier support of $\tilde{F}, \tilde{G}$ the multipliers m_* and $\eta_{\leq -u} \otimes \eta_{\leq -du}$ are both equal to 1. Since $\mathcal{F}_{\mathbb{Z}} \tilde{F}$ vanishes on $\mathcal{M}_{\leq l_1 - 1, \leq -\operatorname{Log} N + l_1 - 1}$ and $\mathcal{F}_{\mathbb{Z}} \tilde{G}$

vanishes on the major arcs $\mathcal{M}_{\leq l_2-1,\leq -d\operatorname{Log} N+dl_2-d}$, we see from Theorem 5.12 (and (1.6)) that

$$\|\tilde{A}_{2^u}(\tilde{F},\tilde{G})\|_{\ell^p(\mathbb{Z})} \lesssim_{C_3} 2^{-cl\mathbb{1}_{p_1=p_2=2}}.$$

By the triangle inequality, it thus suffices to show that

$$\|\mathrm{B}^{l_1,l_2,m_{\hat{\mathbb{Z}}}}_{(1-2\tilde{m}_{2^u,\mathbb{R}})m_*}(\tilde{F},\tilde{G})\|_{\ell^p(\mathbb{Z})} \lesssim_{C_3} 2^{O(2^{\rho l})-u}.$$

Applying Lemma 4.20(ii) (and Lemma 5.2(iii)) with $r_1 = 2^{-2u}$ and $r_2 = 2^{-2du}$, it suffices to show that

$$\frac{\partial^{j_1}}{\partial\xi_1^{j_1}}\frac{\partial^{j_2}}{\partial\xi_1^{j_2}}((1-2\tilde{m}_{2^u,\mathbb{R}})m_*)(\xi_1,\xi_2) \lesssim 2^{(2j_1+2dj_2-1)u}$$

for all $\xi_1,\xi_2 \in \mathbb{R}$ and $0 \leq j_1, j_2 \leq 2$. By the product rule and definition of m_*, it suffices to show that

$$\frac{\partial^{j_1}}{\partial\xi_1^{j_1}}\frac{\partial^{j_2}}{\partial\xi_1^{j_2}}(1-2\tilde{m}_{2^u,\mathbb{R}})(\xi_1,\xi_2) \lesssim 2^{(2j_1+2dj_2-1)u}$$

when $\xi_1 = O(2^{-2u})$, $\xi_2 = O(2^{-2du})$, and $0 \leq j_1, j_2 \leq 2$. But from (7.9) one has

$$\begin{aligned}1-2\tilde{m}_{2^u,\mathbb{R}}(\xi_1,\xi_2) &= 2\int_{1/2}^1 1 - e(2^u t\xi_1 + P(2^u t)\xi_2)\,dt \\ &= -4\pi i\int_0^1\int_{1/2}^1 (2^u\xi_1 + 2^u P'(2^u tt')\xi_2)e(2^u t\xi_1 + P(2^u t)\xi_2)\,dtdt',\end{aligned}$$

so by differentiation under the integral sign and the triangle inequality it suffices to show that

$$\frac{\partial^{j_1}}{\partial\xi_1^{j_1}}\frac{\partial^{j_2}}{\partial\xi_1^{j_2}}(2^u\xi_1 + 2^u P'(2^u tt')\xi_2)e(2^u t\xi_1 + P(2^u t)\xi_2) \lesssim 2^{(2j_1+2dj_2-1)u}$$

uniformly for $t \in [0,1]$, $t' \in [1/2,1]$. But this follows from direct calculation. (In fact, one obtains a slightly stronger bound of $O(2^{(j_1+dj_2-1)u})$ when $j_1 = j_2 = 0$ and $O(2^{(j_1+dj_2)u})$ when $j_1 + j_2 > 0$.) □

In view of this lemma, it now suffices to establish the bounds

$$\|\mathrm{T}^{l_1}_{\varphi_*}\|_{\ell^q(\mathbb{Z})\to\ell^q(\mathbb{Z})}, \|\mathrm{T}^{l_2}_{\tilde{\varphi}_*}\|_{\ell^q(\mathbb{Z})\to\ell^q(\mathbb{Z})} \lesssim_{C_3,q} \langle\max(l,s_1,s_2)\rangle^{O(1)}$$

for any $1 < q < \infty$. By interpolation, it suffices to achieve this when q is an even integer or the dual of an even integer. Using Theorem 5.7, it suffices to show that

$$\|\mathrm{T}_{\varphi_*}\|_{L^q(\mathbb{R})\to L^q(\mathbb{R})}, \|\mathrm{T}_{\tilde{\varphi}_*}\|_{L^q(\mathbb{R})\to L^q(\mathbb{R})} \lesssim_{C_3,q} \max(1,s_1,s_2)^{O(1)}$$

for all $1 < q < \infty$.

By expanding out (8.5), (7.17), and (7.5) (and treating the $s_2 > -u$, $s_2 = -u$ cases separately), we see that $\tilde{\varphi}_*$ is a shifted Calderón–Zygmund multiplier of the form treated in Theorem B.1, with $A = 2^{-ds_2}$, $\lambda_N = 2^{ds_2}P(Nt)/N^d$,

$K = O(\max(1, s_2))$, and $C = O(1)$. (Note that the hypothesis (7.29) is needed to avoid a divergence at the frequency origin.) The claim for $\mathrm{T}_{\tilde{\varphi}_*}$ then follows from that theorem. The treatment of T_{φ_*} is similar (with s_2 replaced by s_1, $P(Nt)$ replaced by Nt, and d replaced by 1). This concludes the proof of (7.30).

9. The large-scale estimate: exploiting tensor product structure

In this section we establish (7.31). Note from Examples 4.10 and 4.11 that one can factor the bilinear operator $\mathrm{B}_{1\otimes m_{\hat{\mathbb{Z}}}}$ as the tensor product of the identity and the arithmetic averaging operator $A_{\hat{\mathbb{Z}}}$. Thus on the one hand, we can write

$$\mathrm{B}_{1\otimes m_{\hat{\mathbb{Z}}}}(\mathrm{T}_{\varphi_{N,t,j_1}}\otimes 1 F_{\mathbb{A}}, \mathrm{T}_{\tilde{\varphi}_{N,t,j_2}}\otimes 1 G_{\mathbb{A}})$$

as

$$\int_{\hat{\mathbb{Z}}} (\mathrm{T}_{\varphi_{N,t,j_1}}\otimes 1 \tau_{(0,y)} F_{\mathbb{A}})(\mathrm{T}_{\tilde{\varphi}_{N,t,j_2}}\otimes 1 \tau_{(0,P(y))} G_{\mathbb{A}})\, d\mu_{\hat{\mathbb{Z}}}(y), \tag{9.1}$$

where we define the translation operators $\tau_h F(x) := F(x-h)$ for any $F \in L^0(\mathbb{G})$ and $h \in \mathbb{G}$. On the other hand, if we use $F_x \colon y \mapsto F(x,y)$ to denote the slice $F_x \colon \hat{\mathbb{Z}} \to \mathbb{C}$ of a function $F \colon \mathbb{A}_{\mathbb{Z}} \to \mathbb{C}$ at a real number x, we can write the slice

$$\mathrm{B}_{1\otimes m_{\hat{\mathbb{Z}}}}(\mathrm{T}_{\varphi_{N,t,j_1}}\otimes 1 F_{\mathbb{A}}, \mathrm{T}_{\tilde{\varphi}_{N,t,j_2}}\otimes 1 G_{\mathbb{A}})_x$$

as

$$A_{\hat{\mathbb{Z}}}((\mathrm{T}_{\varphi_{N,t,j_1}}\otimes 1 F_{\mathbb{A}})_x, (\mathrm{T}_{\tilde{\varphi}_{N,t,j_2}}\otimes 1 G_{\mathbb{A}})_x). \tag{9.2}$$

We now establish the easier case $(p_1, p_2) \neq (2,2)$, in which we do not need to obtain a gain of the form 2^{-cl}; we will also not need to lose factors of $2^{O(\rho l)}$. As such we will not need to exploit any cancellation in the averaging operator $A_{\hat{\mathbb{Z}}}$, and we can use the formulation (9.1). By the triangle inequality, it thus suffices to show that

$$\|((\mathrm{T}_{\varphi_{N,t,j_1}}\otimes 1 \tilde{F}_{\mathbb{A}})(\mathrm{T}_{\tilde{\varphi}_{N,t,j_2}}\otimes 1 \tilde{G}_{\mathbb{A}}))_{N\in\mathbb{I}_>}\|_{L^p(\mathbb{A}_{\mathbb{Z}};\boldsymbol{V}^r)} \lesssim_{C_3} \langle \max(l, s_1, s_2)\rangle^{O(1)} \|\tilde{F}_{\mathbb{A}}\|_{L^{p_1}(\mathbb{A}_{\mathbb{Z}})} \|\tilde{G}_{\mathbb{A}}\|_{L^{p_2}(\mathbb{A}_{\mathbb{Z}})}$$

for all $\tilde{F}_{\mathbb{A}} \in L^{p_1}(\mathbb{A}_{\mathbb{Z}})$ and $\tilde{G}_{\mathbb{A}} \in L^{p_2}(\mathbb{A}_{\mathbb{Z}})$. There are now no interactions between the different fibers $\mathbb{R}\times\{y\}$, $y \in \hat{\mathbb{Z}}$ of $\mathbb{A}_{\mathbb{Z}}$, and so by Hölder's inequality and the Fubini–Tonelli theorem (or (2.12)), it suffices to prove the continuous bilinear estimate

$$\|((\mathrm{T}_{\varphi_{N,t,j_1}}\tilde{F})(\mathrm{T}_{\tilde{\varphi}_{N,t,j_2}}\tilde{G}))_{N\in\mathbb{I}_>}\|_{L^p(\mathbb{R};\boldsymbol{V}^r)} \lesssim_{C_3} \langle \max(l, s_1, s_2)\rangle^{O(1)} \|\tilde{F}\|_{L^{p_1}(\mathbb{R})} \|\tilde{G}\|_{L^{p_2}(\mathbb{R})}$$

for any $\tilde{F} \in L^{p_1}(\mathbb{R}), \tilde{G} \in L^{p_2}(\mathbb{R})$. By (2.8) and Hölder's inequality it suffices to establish the linear bounds

$$\|(\mathrm{T}_{\varphi_{N,t,j_1}}\tilde{F})_{N\in\mathbb{I}_>}\|_{L^{p_1}(\mathbb{R};\boldsymbol{V}^r)} \lesssim_{C_3} \max(1,s_1)^{O(1)}\|\tilde{F}\|_{L^{p_1}(\mathbb{R})} \tag{9.3}$$

and

$$\|(\mathrm{T}_{\tilde{\varphi}_{N,t,j_2}}\tilde{G})_{N\in\mathbb{I}_>}\|_{L^{p_2}(\mathbb{R};\boldsymbol{V}^r)} \lesssim_{C_3} \max(1,s_2)^{O(1)}\|\tilde{G}\|_{L^{p_2}(\mathbb{R})}. \tag{9.4}$$

We just establish the latter estimate, as the former is similar. First suppose that we are in the high-frequency case $s_2 > -u$. In this case we use (2.7) to replace the $\boldsymbol{V}^r$ norm by an ℓ^2 norm, thus we now wish to show

$$\|(\mathrm{T}_{\tilde{\varphi}_{N,t,j_2}}\tilde{G})_{N\in\mathbb{I}_>}\|_{L^{p_2}(\mathbb{R};\ell^2)} \lesssim_{C_3} \max(1,s_2)^{O(1)}\|\tilde{G}\|_{L^{p_2}(\mathbb{R})}.$$

But as with the arguments at the end of Section 8, the $\tilde{\varphi}_{N,t,j_2}$ form a family of the type considered in Theorem B.1, with $A = 2^{-ds_2}$, $\lambda_N = 2^{ds_2}P(Nt)/N^d$, $K = O(\max(1,s_2))$, and $C = O(1)$. The claim now follows from the shifted square function estimate proven in that theorem.

Now suppose we are in the low-frequency case $s_2 = -u$, which means that $j_2 = 0, 1$ by the hypothesis (7.29). If $j_2 = 1$, then $\tilde{\varphi}_{N,t,j_2}$ vanishes at the origin and we can repeat the arguments from the high-frequency case. If $j_2 = 0$, then $\tilde{\varphi}_{N,t,j_2} = \tilde{\varphi}_{N,t}$ no longer vanishes at the origin, but the difference $\tilde{\varphi}_{N,t} - \tilde{\varphi}_N$ does, and we can again use the high-frequency arguments to conclude. By the triangle inequality, it now suffices to show that

$$\|(\mathrm{T}_{\tilde{\varphi}_N}\tilde{G})_{N\in\mathbb{I}_>}\|_{L^{p_2}(\mathbb{R};\boldsymbol{V}^r)} \lesssim_{C_3} \|\tilde{G}\|_{L^{p_2}(\mathbb{R})}.$$

But this follows from Lépingle's inequality and a standard square function argument. (See [50, Th. 1.1], with the square function argument contained in [50, Lemma 3.2].)

This completes the proof of the $(p_1,p_2) \neq (2,2)$ case of (7.31). Now we turn to the $(p_1,p_2) = (2,2)$ case, so that $p = 1$. We begin with a general variational inequality:

LEMMA 9.5 (Interchanging variational and Lebesgue norms). *Let X be a measure space, and let $1 \leq R < r \leq \infty$. Then for any $f_1,\dots,f_K \in L^r(X)$, one has*

$$\|(f_k)_{k\in[K]}\|_{L^r(X;\boldsymbol{V}^r)} \lesssim_{r,R} \|(f_k)_{k\in[K]}\|_{\boldsymbol{V}^R([K];L^r(X))}.$$

Proof. We allow implied constants to depend on r, R. Since

$$\|(f_k)_{k\in[K]}\|_{L^r(X;\boldsymbol{V}^r)} \lesssim \|(f_k)_{k\in[K]}\|_{L^r(X;V^r)} + \|f_1\|_{L^r(X)},$$

it suffices to establish the seminorm version

$$\|(f_k)_{k\in[K]}\|_{L^r(X;V^r)} \lesssim_{r,R} \|(f_k)_{k\in[K]}\|_{V^R([K];L^r(X))}$$

of the inequality.

We can assume that f_k is not almost everywhere equal to f_{k-1} for any $1 < k \leq K$, since otherwise we could concatenate the two indices $k, k-1$ together. We normalize

$$\|(f_k)_{k\in[K]}\|^R_{V^R([K];L^r(X))} = 1.$$

Then we can define a non-decreasing function $a\colon [K] \to [0,1]$ by the formula

$$a(K') := \|(f_k)_{k\in[K']}\|^R_{V^R([K'];L^r(X))}$$

for any $K' \in [K]$. From (2.4), we have the Hölder type bound

$$\|f_{K_1} - f_{K_2}\|_{L^r(X)} \leq (a(K_1) - a(K_2))^{1/R} \tag{9.6}$$

whenever $1 \leq K_2 \leq K_1 \leq K$. In particular, because we assumed f_k not equal almost everywhere to f_{k-1}, we see that a is strictly increasing.

For any $x \in X$, let μ_x be the absolutely continuous complex measure on $[0,1]$ defined by

$$\mu_x(E) := \sum_{2\leq k\leq K : a(k)\in E} \frac{|E \cap [a(k-1), a(k)]|}{|[a(k-1), a(k)]|}(f_k(x) - f_{k-1}(x)).$$

Then we have

$$f_{K_1}(x) - f_{K_2}(x) = \mu_x([a(K_2), a(K_1)])$$

whenever $1 \leq K_2 \leq K_1 \leq K$. Also from (9.6) and telescoping series (and the hypothesis $R \geq 1$), we observe the Hölder bound

$$\|\mu_x([s,t])\|_{L^r(X)} \lesssim (t-s)^{1/R} \tag{9.7}$$

for any $0 \leq s \leq t \leq 1$.

Using dyadic decomposition as in the proof of Lemma 8.1 (or [67, Lemma 2.5, pp. 534]), we have

$$\|(f_k(x))_{k\in[K]}\|_{V^r} \lesssim \sum_{m=0}^{\infty} \big\|(\mu_x([(j-1)2^{-m}, j2^{-m}))_{j\in[2^m]}\big\|_{\ell^r}$$

and hence by the Fubini–Tonelli theorem and the triangle inequality,

$$\|(f_k)_{k\in[K]}\|_{L^r(X;V^r)} \lesssim \sum_{m=0}^{\infty} \Big\|(\|\mu_x([(j-1)2^{-m}, j2^{-m}))\|_{L^r(X)})_{j\in[2^m]}\Big\|_{\ell^r}.$$

Applying (9.7), the right-hand side is

$$\lesssim \sum_{m=0}^{\infty} 2^{m/r} 2^{-m/R};$$

since $R < r$, this quantity is $O(1)$, and the claim follows. □

We can apply this lemma to bilinear operators:

COROLLARY 9.8 (Interchanging variational and Lebesgue norms, II). *Let V, W be normed vector spaces, let $K \in \mathbb{Z}_+$, and for each $k \in [K]$, let $f_k \in V, g_k \in W$. Let $1 \leq R < r \leq \infty$, and let $B : V \times W \to L^r(X)$ be a bilinear map to $L^r(X)$ for some measure space X. Then*

$$\|(B(f_k, g_k))_{k\in[K]}\|_{L^r(X;\boldsymbol{V}^r)} \lesssim_{r,R} \|B\|_{V\times W\to L^r(X)} \|(f_k)_{k\in[K]}\|_{\boldsymbol{V}^R([K];V)} \|(g_k)_{k\in[K]}\|_{\boldsymbol{V}^R([K];W)}.$$

Proof. We allow all implied constants to depend on r, R. We may normalize

$$\|B\|_{V\times W\to L^r(X)} = \|(f_k)_{k\in[K]}\|_{\boldsymbol{V}^R([K];V)} = \|(g_k)_{k\in[K]}\|_{\boldsymbol{V}^R([K];W)} = 1.$$

In particular, the product sequence $(f_k, g_k) \in V \times W$, $k \in [K]$ obeys the variational norm bound

$$\|(f_k, g_k)_{k\in[K]}\|_{\boldsymbol{V}^R([K];V\times W)} \lesssim 1.$$

By Lemma 9.5, it suffices to show that

$$\|B(f_k, g_k))_{k\in[K]}\|_{\boldsymbol{V}^R([K];L^r(X))} \lesssim 1.$$

On the ball of radius $O(1)$ in $V \times W$, the (nonlinear) map $(f, g) \mapsto B(f, g)$ is Lipschitz continuous into $L^r(X)$ with Lipschitz constant $O(1)$, and the claim follows from (2.5). □

We apply this lemma to the problem of establishing (7.31) in the $p_1 = p_2 = 2$ case. In the next section we establish the following arithmetic variant of Theorem 5.12:

THEOREM 9.9 (Arithmetic bilinear estimate). *Let $l \in \mathbb{N}$, and let $f, g \in L^2(\hat{\mathbb{Z}})$ obey one of the following hypotheses*:

(i) *$\mathcal{F}_{\hat{\mathbb{Z}}} f$ vanishes on $(\mathbb{Q}/\mathbb{Z})_{\leq l}$;*
(ii) *$\mathcal{F}_{\hat{\mathbb{Z}}} g$ vanishes on $(\mathbb{Q}/\mathbb{Z})_{\leq l}$.*

Then for any $1 \leq r < \frac{2d}{d-1}$, one has

$$\|A_{\hat{\mathbb{Z}}}(f, g)\|_{L^r(\hat{\mathbb{Z}})} \lesssim_{C_3,r} 2^{-c_r l} \|f\|_{L^2(\hat{\mathbb{Z}})} \|g\|_{L^2(\hat{\mathbb{Z}})}.$$

(*Recall our conventions that $c_r > 0$ denotes a constant that can depend on d, r.*)

The key point here is that the exponent r in Theorem 9.9 is allowed to be slightly larger than 2.

To prove (7.31) for $r > 2$, we use the slice formulation (9.2). It suffices by monotonicity of $\boldsymbol{V}^r$ norms to work in the range $2 < r < \frac{2d}{d-1}$. From (7.18) and (7.19), we see that every slice $(F_{\mathbb{A}})_x$, $(G_{\mathbb{A}})_x$ of $F_{\mathbb{A}}, G_{\mathbb{A}}$ take values in the finite-dimensional vector spaces $L^2(\hat{\mathbb{Z}})^{(\mathbb{Q}/\mathbb{Z})_{l_1}}, L^2(\hat{\mathbb{Z}})^{(\mathbb{Q}/\mathbb{Z})_{l_2}}$ respectively, and hence so do $\mathrm{T}_{\varphi_{N,t,j_1}\otimes 1}\vec{F}, \mathrm{T}_{\tilde{\varphi}_{N,t,j_2}\otimes 1}\vec{G}$ for any N. By Theorem 9.9, the operator norm of $A_{\hat{\mathbb{Z}}} : L^2(\hat{\mathbb{Z}})^{(\mathbb{Q}/\mathbb{Z})_{l_1}} \times L^2(\hat{\mathbb{Z}})^{(\mathbb{Q}/\mathbb{Z})_{l_2}} \to L^r(\mathbb{Z}/Q\mathbb{Z})$ is $O_{C_3}(2^{-cl})$. Applying

Hölder's inequality to bound the $L^1(\mathbb{A}_{\mathbb{Z}})$ norm by the $L^r(\mathbb{A}_{\mathbb{Z}})$ norm, followed Corollary 9.8 for some $2 < R < r$, then Cauchy–Schwarz, we conclude[9] that

$$\begin{aligned}\|(A_{\hat{\mathbb{Z}}}(\mathrm{T}_{\varphi_{N,t,j_1}}\otimes 1(F_{\mathbb{A}})_x, \mathrm{T}_{\tilde{\varphi}_{N,t,j_2}}\otimes 1(G_{\mathbb{A}})_x))_{N\in\mathbb{I}_>}\|_{L^1(\mathbb{R};L^1(\hat{\mathbb{Z}};\boldsymbol{V}^r))}\\ \lesssim_{C_3} 2^{-cl}\|(\mathrm{T}_{\varphi_{N,t,j_1}}\otimes 1(F_{\mathbb{A}})_x)_{N\in\mathbb{I}_>}\|_{L^2(\mathbb{R};\boldsymbol{V}^R(\mathbb{I}_>;L^2(\hat{\mathbb{Z}})))}\\ \times \|(\mathrm{T}_{\tilde{\varphi}_{N,t,j_2}}\otimes 1(G_{\mathbb{A}})_x)_{N\in\mathbb{I}_>}\|_{L^2(\mathbb{R};\boldsymbol{V}^R(\mathbb{I}_>;L^2(\hat{\mathbb{Z}})))},\end{aligned}$$

where we view x as a variable of integration in $\mathbb{R}$. It thus suffices to establish the bounds

$$\|(\mathrm{T}_{\varphi_{N,t,j_1}}\otimes 1\vec{F})_{N\in\mathbb{I}_>}\|_{L^2(\mathbb{R};\boldsymbol{V}^R(\mathbb{I}_>;L^2(\hat{\mathbb{Z}})))} \lesssim_{C_3} \max(1,s_1)^{O(1)}\|\vec{F}\|_{L^2(\mathbb{R};L^2(\hat{\mathbb{Z}}))}$$

and

$$\|(\mathrm{T}_{\tilde{\varphi}_{N,t,j_2}}\otimes 1\vec{G})_{N\in\mathbb{I}_>}\|_{L^2(\mathbb{R};\boldsymbol{V}^R(\mathbb{I}_>;L^2(\hat{\mathbb{Z}})))} \lesssim_{C_3} \max(1,s_2)^{O(1)}\|\vec{G}\|_{L^2(\mathbb{R};L^2(\hat{\mathbb{Z}}))}$$

for any vector-valued functions $\vec{F}, \vec{G} \in L^2(\mathbb{R}; L^2(\hat{\mathbb{Z}}))$. But these are simply vector-valued versions of (9.3) and (9.4), and they are proven in exactly the same fashion (since all of the tools used in the proof extend to the vector-valued setting); in particular, the vector-valued version of Lépingle's inequality was established in [69, Th. 3.1, pp. 810], and all linear L^p estimates extend to the vector-valued setting by the Marcinkiewicz–Zygmund inequality. One may first wish to approximate $L^2(\hat{\mathbb{Z}})$ by a finite dimensional Hilbert space to avoid technicalities. This will conclude the proof of (7.25) (and thus Theorem 1.17), once we establish Theorem 9.9. This is the purpose of the next section.

10. **Arithmetic bilinear estimates**

We now prove Theorem 9.9. It may be worth mentioning that the adelic viewpoint is not strictly necessary here and one could replace the profinite integers $\hat{\mathbb{Z}}$ here with $\mathbb{Z}/Q\mathbb{Z}$. But then one needs to check that none of the bounds lose any factor of Q (or even $\log Q$) as this would be fatal to the argument. From this point of view, the adelic formalism is cleaner and automatically handles uniformity in the Q parameter. We begin with the $r = 1$ case, which is a limiting case of Theorem 5.12 in which the continuous aspect of that theorem degenerates completely, leaving only the arithmetic aspect:

PROPOSITION 10.1. *Theorem 9.9 holds when $r = 1$.*

[9] Strictly speaking, our definitions and arguments are not justified here because the vector spaces $L^2(\hat{\mathbb{Z}})$, $L^1(\hat{\mathbb{Z}};\boldsymbol{V}^r)$ are infinite-dimensional. However, one can approximate $\hat{\mathbb{Z}}$ by finite cyclic groups $\mathbb{Z}/Q\mathbb{Z}$ to make these spaces finite-dimensional and then take limits to avoid this difficulty; indeed, given the definitions of $F_{\mathbb{A}}, G_{\mathbb{A}}$, we can just work with a single large but fixed Q. Alternatively one can extend many of the previous vector-valued definitions to separable Banach spaces. We leave the details to the interested reader.

We remark that when q is a prime this result is essentially contained in [15] (when $P(\mathrm{n}) = \mathrm{n}^2$) and [72] (in the general case); see [27] for the strongest current values for the constant c.

Proof. For the sake of exposition, we assume that hypothesis (i) of Theorem 9.9 holds; for the case when hypothesis (ii) is assumed, one proceeds similarly. By a limiting argument we may assume that the functions F, G on $\hat{\mathbb{Z}}$ factor through a finite quotient $\mathbb{Z}/Q\mathbb{Z}$, in which case the task is to show that

$$\|A_{\mathbb{Z}/Q\mathbb{Z}}(f,g)\|_{L^1(\mathbb{Z}/Q\mathbb{Z})} \lesssim_{C_3} 2^{-cl}\|f\|_{L^2(\mathbb{Z}/Q\mathbb{Z})}\|g\|_{L^2(\mathbb{Z}/Q\mathbb{Z})}$$

assuming that $\mathcal{F}_{\mathbb{Z}/Q\mathbb{Z}}f$ vanishes on $(\mathbb{Q}/\mathbb{Z})_{\leq l} \cap (\frac{1}{Q}\mathbb{Z}/\mathbb{Z})$.

Let N be a large natural number (which we will eventually send to infinity), and let R be an extremely large real number (which we will also send to infinity, before sending N to infinity). In particular, one should think of N, R as being large compared to l, Q. We define the functions $f_R, g_R \in \boldsymbol{S}(\mathbb{Z})$ by the formulae

$$f_R(n) \coloneqq \frac{1}{\sqrt{R}}\psi(n/R)f(n \bmod Q),$$
$$g_R(n) \coloneqq \frac{1}{\sqrt{R}}\psi(n/R)g(n \bmod Q),$$

where $\psi \in \boldsymbol{S}(\mathbb{R})$ is a real even function with $\|\psi\|_{L^2(\mathbb{R})} = 1$ whose Fourier transform is supported on $[-1,1]$. Clearly $f_R \in L^2(\mathbb{Z})$ has Fourier transform supported on the set $\pi([-1/R, 1/R] \times \{\alpha \in \frac{1}{Q}\mathbb{Z}/\mathbb{Z} \colon \hat{f}(\alpha) \neq 0\})$. From the hypothesis (i), we see that if N, R is sufficiently large (depending on Q, l), this union of arcs is disjoint from all of the arcs in $\mathcal{M}_{\leq l, \leq -\operatorname{Log} N + l}$ (because the frequencies α with $\hat{f}(\alpha) \neq 0$ have a non-zero separation from the frequencies $(\mathbb{Q}/\mathbb{Z})_{\leq l}$). By Theorem 5.12, we conclude for N, R sufficiently large that

$$\|A_{N,\mathbb{Z}}(f_R, g_R)\|_{L^1(\mathbb{Z})} \lesssim_{C_1} (2^{-cl} + \operatorname{Log} N^{-cC_1})\|f_R\|_{L^2(\mathbb{Z})}\|g_R\|_{L^2(\mathbb{Z})}.$$

From the Riemann integrability of $|\psi|^2$, it is easy to see that

$$\lim_{R\to\infty} \|f_R\|_{L^2(\mathbb{Z})} = \|f\|_{L^2(\mathbb{Z}/Q\mathbb{Z})}$$

and similarly

$$\lim_{R\to\infty} \|g_R\|_{L^2(\mathbb{Z})} = \|g\|_{L^2(\mathbb{Z}/Q\mathbb{Z})};$$

hence

$$\limsup_{N\to\infty} \limsup_{R\to\infty} \|A_{N,\mathbb{Z}}(f_R, g_R)\|_{L^1(\mathbb{Z})} \lesssim 2^{-cl}\|f\|_{L^2(\mathbb{Z}/Q\mathbb{Z})}\|g\|_{L^2(\mathbb{Z}/Q\mathbb{Z})}.$$

For any N, R, the Schwartz function nature of ψ readily gives the asymptotic

$$A_{N,\mathbb{Z}}(f_R, g_R)(n) = \frac{1}{R}|\psi(n/R)|^2 A_{\mathbb{Z}/Q\mathbb{Z}}(f,g)(n \bmod Q) + O_{N,Q}\big(R^{-2}\langle n/R\rangle^{-10}\big).$$

Thus by the Riemann integrability of $|\psi|^2$, we obtain

$$\limsup_{R\to\infty} \|A_{N,\mathbb{Z}}(f_R, g_R)\|_{L^1(\mathbb{Z})} = \|A_{\mathbb{Z}/Q\mathbb{Z}}(f,g)\|_{L^1(\mathbb{Z}/Q\mathbb{Z})}.$$

Taking limits as $N \to \infty$, we then have

$$\limsup_{N\to\infty} \limsup_{R\to\infty} \|A_{N,\mathbb{Z}}(f_R, g_R)\|_{L^1(\mathbb{Z})} = \|A_{\mathbb{Z}/Q\mathbb{Z}}(f,g)\|_{L^1(\mathbb{Z}/Q\mathbb{Z})}$$

and the claim follows. □

By interpolation with Proposition 10.1, we see that to establish the remaining cases of Theorem 9.9, it will suffice to establish the bound

$$\|A_{\hat{\mathbb{Z}}}\|_{L^2(\hat{\mathbb{Z}})\times L^2(\hat{\mathbb{Z}})\to L^q(\hat{\mathbb{Z}})} \lesssim_q 1 \tag{10.2}$$

for all $1 \le q < \frac{2d}{d-1}$. Approximating $\hat{\mathbb{Z}}$ by the product of finitely many of the p-adic groups $\mathbb{Z}_p$, it suffices by limiting arguments to show that

$$\|A_{\prod_{p\in S}\mathbb{Z}_p}\|_{L^2(\prod_{p\in S}\mathbb{Z}_p)\times L^2(\prod_{p\in S}\mathbb{Z}_p)\to L^q(\prod_{p\in S}\mathbb{Z}_p)} \lesssim_q 1$$

whenever S is a finite set of primes. From Examples 4.10 and 4.11, we see that the bilinear operator $A_{\prod_{p\in S}\mathbb{Z}_p}$ is the tensor product of the individual operators $A_{\mathbb{Z}_p}$, so by (2.12) we may factor the operator norm as

$$\|A_{\prod_{p\in S}\mathbb{Z}_p}\|_{L^2(\prod_{p\in S}\mathbb{Z}_p)\times L^2(\prod_{p\in S}\mathbb{Z}_p)\to L^q(\prod_{p\in S}\mathbb{Z}_p)} = \prod_{p\in S} \|A_{\mathbb{Z}_p}\|_{L^2(\mathbb{Z}_p)\times L^2(\mathbb{Z}_p)\to L^q(\mathbb{Z}_p)}.$$

Thus it will suffice to establish the bound

$$\|A_{\mathbb{Z}_p}\|_{L^2(\mathbb{Z}_p)\times L^2(\mathbb{Z}_p)\to L^q(\mathbb{Z}_p)} \lesssim_q 1 \tag{10.3}$$

for all primes p, together with the improvement

$$\|A_{\mathbb{Z}_p}\|_{L^2(\mathbb{Z}_p)\times L^2(\mathbb{Z}_p)\to L^q(\mathbb{Z}_p)} \le 1 \tag{10.4}$$

whenever p is sufficiently large depending on q.

We begin with (10.3). By bilinear interpolation, it suffices to establish the bounds

$$\|A_{\mathbb{Z}_p}\|_{L^1(\mathbb{Z}_p)\times L^\infty(\mathbb{Z}_p)\to L^\infty(\mathbb{Z}_p)} \le 1 \tag{10.5}$$

and

$$\|A_{\mathbb{Z}_p}\|_{L^\infty(\mathbb{Z}_p)\times L^1(\mathbb{Z}_p)\to L^s(\mathbb{Z}_p)} \lesssim_s 1 \tag{10.6}$$

for all $1 \le s < \frac{d}{d-1}$. The estimate (10.5) is immediate from the pointwise inequality

$$|A_{\mathbb{Z}_p}(f,g)| \le A^{\mathrm{n}}_{\mathbb{Z}_p}(|f|)\|g\|_{L^\infty(\mathbb{Z}_p)}.$$

To prove (10.6), we similarly use the pointwise inequality

$$|A_{\mathbb{Z}_p}(f,g)| \le A^{P(\mathrm{n})}_{\mathbb{Z}_p}(|g|)\|f\|_{L^\infty(\mathbb{Z}_p)},$$

so it suffices to show the linear L^p improving bound

$$\|A_{\mathbb{Z}_p}^{P(\mathrm{n})}\|_{L^1(\mathbb{Z}_p)\to L^s(\mathbb{Z}_p)} \lesssim_s 1$$

for $1 \le s < \frac{d}{d-1}$. By a limiting argument, it suffices to show that

$$\|A_{\mathbb{Z}/p^j\mathbb{Z}}^{P(\mathrm{n})}\|_{L^1(\mathbb{Z}/p^j\mathbb{Z})\to L^s(\mathbb{Z}/p^j\mathbb{Z})} \lesssim_s 1$$

for all $j \in \mathbb{N}$. By Minkowski's inequality, it suffices to show that the counting function $h\colon \mathbb{Z}/p^j\mathbb{Z} \to \mathbb{N}$ defined by

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

has an $L^s(\mathbb{Z}/p^j\mathbb{Z})$ norm of $O_s(1)$. But this follows from Corollary C.2 in the appendix. This concludes the proof of (10.3). We remark that this argument in fact yields a weak-type endpoint for (10.3), but it is not clear to us how to use this to obtain a corresponding weak-type endpoint for (10.2) as the weak L^p spaces do not interact well with tensor products. In any event, for our application, any exponent q greater than 2 would suffice, so endpoint estimates are not needed.

Now we prove (10.4). By Hölder's inequality, we may take $2 < q < \frac{2d}{d-1}$. We let l be a large number (depending on q, P) to be chosen later, and then assume that $p \in \mathbb{P}$ is a prime that is sufficiently large depending on l, q, P. From Proposition 10.1, we then see that

$$\|A_{\mathbb{Z}_p}(f, g_0)\|_{L^1(\mathbb{Z}_p)} \lesssim_{C_3} 2^{-cl}\|f\|_{L^2(\mathbb{Z}_p)}\|g_0\|_{L^2(\mathbb{Z}_p)},$$

whenever $f, g_0 \in L^2(\mathbb{Z}_p)$ with g_0 of mean zero, since for p large enough, the only element of $\mathbb{Z}_p^*$ of height at most 2^l is the origin.

Interpolating this bound with (10.3) (for a slightly larger choice of q), we conclude that

$$\|A_{\mathbb{Z}_p}(f, g_0)\|_{L^q(\mathbb{Z}_p)} \lesssim_{q,C_3} 2^{-c_q l}\|f\|_{L^2(\mathbb{Z}_p)}\|g_0\|_{L^2(\mathbb{Z}_p)}. \tag{10.7}$$

(Recall our conventions that $c_q > 0$ denotes a constant that can depend on d, q and varies from line to line.)

Let $f, g \in L^2(\mathbb{Z}_p)$ with $\|f\|_{L^2(\mathbb{Z}_p)} = \|g\|_{L^2(\mathbb{Z}_p)} = 1$. It will suffice to show that

$$\mathbb{E}_{n\in\mathbb{Z}_p}|A_{\mathbb{Z}_p}(f, g)(n)|^q \le 1.$$

Since $|A_{\mathbb{Z}_p}(f, g)| \le A_{\mathbb{Z}_p}(|f|, |g|)$, we may assume without loss of generality that f, g are non-negative. We split $f = a + f_0$ and $g = b + g_0$, where

$$a := \mathbb{E}_{n\in\mathbb{Z}_p} f(n) \quad \text{and} \quad b := \mathbb{E}_{n\in\mathbb{Z}_p} g(n)$$

are the means of f, g, and $f_0 := f - a$, $g_0 := g - b$ are the mean zero components. If we define the "energies"

$$E_f := \|f_0\|_{L^2(\mathbb{Z}_p)}^2 \quad \text{and} \quad E_g := \|g_0\|_{L^2(\mathbb{Z}_p)}^2,$$

then from Pythagoras' theorem we have $0 \leq E_f, E_g \leq 1$ and

$$|a| = (1 - E_f)^{1/2} \quad \text{and} \quad |b| = (1 - E_g)^{1/2}. \tag{10.8}$$

A short calculation shows that

$$A_{\mathbb{Z}_p}(a, b) = ab \quad \text{and} \quad A_{\mathbb{Z}_p}(f_0, b) = 0$$

and hence

$$A_{\mathbb{Z}_p}(f, g) = ab + A_{\mathbb{Z}_p}(f, g_0).$$

Since the function $x \mapsto |x|^q$ is continuously twice differentiable, Taylor expansion yields the pointwise bound

$$|A_{\mathbb{Z}_p}(f, g)|^q = |ab|^q + q|ab|^{q-1} A_{\mathbb{Z}_p}(f, g_0) + O_q(|A_{\mathbb{Z}_p}(f, g_0)|^2 + |A_{\mathbb{Z}_p}(f, g_0)|^q).$$

Since $A_{\mathbb{Z}_p}(a, g_0)$ has mean zero, we have

$$\begin{aligned} \mathbb{E}_{n \in \mathbb{Z}_p} A_{\mathbb{Z}_p}(f, g_0)(n) &= \mathbb{E}_{n \in \mathbb{Z}_p} A_{\mathbb{Z}_p}(f_0, g_0)(n) \\ &\leq \|A_{\mathbb{Z}_p}(f_0, g_0)\|_{L^1(\mathbb{Z}_p)} \end{aligned}$$

and thus (since $|a|, |b| \leq 1$ and $q \geq 2$)

$$\begin{aligned} \|A_{\mathbb{Z}_p}(f, g)\|_{L^q(\mathbb{Z}_p)}^q \leq |ab|^2 &+ O_q(\|A_{\mathbb{Z}_p}(f_0, g_0)\|_{L^1(\mathbb{Z}_p)} \\ &+ \|A_{\mathbb{Z}_p}(f, g_0)\|_{L^2(\mathbb{Z}_p)}^2 + \|A_{\mathbb{Z}_p}(f, g_0)\|_{L^q(\mathbb{Z}_p)}^q). \end{aligned}$$

From (10.8) and (10.7), the L^2 boundedness of f, f_0, g_0, and Hölder's inequality, we conclude

$$\|A_{\mathbb{Z}_p}(f, g)\|_{L^q(\mathbb{Z}_p)}^q \leq (1 - E_f)(1 - E_g) + O_{q, C_3}(2^{-c_q l}(E_f^{1/2} E_g^{1/2} + E_g)).$$

Since $E_f E_g \leq \min(E_f, E_g) \leq \frac{E_f + E_g}{2}$, one has

$$(1 - E_f)(1 - E_g) \leq 1 - \frac{E_f + E_g}{2};$$

since $E_f^{1/2} E_g^{1/2} = O(E_f + E_g)$, the claim follows by choosing l large enough depending on q, C_3. This proves (10.6), and thus Theorem 9.9.

The proof of Theorem 1.17 is (finally!) complete.

11. **Breaking duality**

In this section we extend Theorem 1.17 to certain cases in which $p < 1$. Throughout this section $P \in \mathbb{Z}[\mathrm{n}]$ is a polynomial of degree $d \geq 2$.

We begin with the following expansion of the range of applicability of (1.6) for these averages.

LEMMA 11.1 (Single scale estimate below ℓ^1). *Let $1 < p_1, p_2 < \infty$ obey the constraints*

$$\frac{1}{p_1} + \frac{2}{p_2}, \quad \frac{2}{p_1} + \frac{1}{p_2} < 2 \tag{11.2}$$

when $d = 2$, or

$$\frac{d^2+d-1}{p_1} + \frac{d^2+d+1}{p_2}, \quad \frac{d^2+d+1}{p_1} + \frac{d^2+d-1}{p_2} < d^2+d+1$$

when $d \geq 3$. Then for any measure-preserving system (X, μ, T), one has

$$\|A_N^{\mathrm{n},P(\mathrm{n})}(f,g)\|_{L^p(X)} \lesssim_{p_1,p_2,P} \|f\|_{L^{p_1}(X)}\|g\|_{L^{p_2}(X)}$$

for all $N \geq 1$, $f \in L^{p_1}(X)$, $g \in L^{p_2}(X)$, where $\frac{1}{p} = \frac{1}{p_1} + \frac{1}{p_2}$; similarly with $A_N^{\mathrm{n},P(\mathrm{n})}$ replaced by $\tilde{A}_N^{\mathrm{n},P(\mathrm{n})}$.

We remark that if [41, Conj. 1.5] holds, the condition should be able to be relaxed to

$$\frac{d-1}{p_1} + \frac{d}{p_2}, \quad \frac{d}{p_1} + \frac{d-1}{p_2} < d,$$

bringing it in line with (11.2).

Proof. From the pointwise bound $|\tilde{A}_N^{\mathrm{n},P(\mathrm{n})}(f,g)| \leq A_N^{\mathrm{n},P(\mathrm{n})}(|f|,|g|)$, it suffices to establish the claim for $A_N^{\mathrm{n},P(\mathrm{n})}$. We may assume that $p < 1$ since the claim follows from (1.6) otherwise. By the Calderón transference principle, it suffices to establish this bound for the case of the integer shift $(\mathbb{Z}, \mu_{\mathbb{Z}}, T_{\mathbb{Z}})$. Noting the pointwise bound

$$|A_N^{\mathrm{n},P(\mathrm{n})}(f,g)(x)| \leq \sum_{I \in \mathcal{I}} \mathbb{1}_I(x) A_N^{\mathrm{n},P(\mathrm{n})}(\mathbb{1}_I|f|, \mathbb{1}_I|g|)(x),$$

where I ranges over a collection $\mathcal{I}$ of intervals of length $O_P(N^d)$ and overlap $O_P(1)$, it suffices to establish the claimed bound when f, g are supported in a single one of these intervals I, that is to say,

$$\|A_N^{\mathrm{n},P(\mathrm{n})}(f,g)\|_{\ell^p(\mathbb{Z})} \lesssim_{p_1,p_2,P} \|f\|_{\ell^{p_1}(I)}\|g\|_{\ell^{p_2}(I)}.$$

As $A_N^{\mathrm{n},P(\mathrm{n})}(f,g)$ is supported in an interval of length $O_P(N^d)$, we have from Hölder's inequality and the hypothesis $p < 1$ that

$$\|A_N^{\mathrm{n},P(\mathrm{n})}(f,g)\|_{\ell^p(\mathbb{Z})} \lesssim_{p_1,p_2,P} N^{d(\frac{1}{p}-1)}\|A_N^{\mathrm{n},P(\mathrm{n})}(f,g)\|_{\ell^1(\mathbb{Z})}.$$

From the triangle inequality and the Fubini–Tonelli theorem, one has

$$\|A_N^{\mathrm{n},P(\mathrm{n})}(f,g)\|_{\ell^1(\mathbb{Z})} \leq \sum_{x \in \mathbb{Z}} |f|(x) A^{P(\mathrm{n})-\mathrm{n}}|g|(x)$$

(cf. (6.1)), so by Hölder's inequality, it suffices to establish the bound

$$\|A^{P(\mathrm{n})-\mathrm{n}}g\|_{\ell^{p_2'}(\mathbb{Z})} \lesssim_{p_1,p_2,P} N^{d(\frac{1}{p_2'}-\frac{1}{p_1})}\|g\|_{\ell^{p_1}(\mathbb{Z})}$$

for any $g \in \ell^{p_1}(\mathbb{Z})$. But this follows from the results of [41] (cf. Proposition 6.21). □

As remarked in the proof of Proposition 6.21, one expects the range of p_1, p_2 to be improvable here, at least in the case $d \geq 3$. We remark that the same argument allows one to break duality in (1.6) (that is to say, obtain (1.6) for at least some ranges of exponents $p_1, \dots, p_k$ with $\frac{1}{p_1} + \cdots + \frac{1}{p_k} > 1$) for any average $A_N^{P_1(\mathrm{n}),\dots,P_k(\mathrm{n})}$ (or $\tilde{A}_N^{P_1(\mathrm{n}),\dots,P_k(\mathrm{n})}$) in which all the P_i have degree at most d, with at least one of the differences $P_i - P_j$ having degree exactly d, for some $d \geq 2$; we leave the details to the interested reader.

Now we can obtain norm convergence results with an explicit range of p_1, p_2.

Corollary 11.3 (Breaking duality for the mean ergodic theorem). *Let (X, μ, T) be a measure-preserving system with X of finite measure, and let $P(\mathrm{n}) \in \mathbb{Z}[\mathrm{n}]$ have degree $d \geq 2$. If p_1, p_2, p obey the hypotheses in Lemma 11.1, then the averages $A_N^{\mathrm{n}, P(\mathrm{n})}(f, g)$ converge in $L^p(X)$ norm for all $f \in L^{p_1}(X)$, $g \in L^{p_2}(X)$.*

Proof. By Theorem 1.17(i) and Hölder's inequality (using the finite measure hypothesis), the claim already holds for (say) $f, g \in L^\infty(X)$. The claim now follows from Lemma 11.1 and the usual limiting argument (which is still valid in the quasinormed space $L^p(X)$). □

For the remaining components of Theorem 1.17, we can similarly break duality, albeit with a much poorer range of exponents:

Proposition 11.4 (Breaking duality for all the ergodic theorems). *Let $P(\mathrm{n}) \in \mathbb{Z}[\mathrm{n}]$ have degree $d \geq 2$, and let $\varepsilon > 0$. If $(\frac{1}{p_1}, \frac{1}{p_2})$ is in a sufficiently small neighborhood of $(\frac{1}{2}, \frac{1}{2})$ (where the neighborhood depends only on d,ε), and $\frac{1}{p} := \frac{1}{p_1} + \frac{1}{p_2}$, then the conclusions* (i)–(iv) *of Theorem 1.17 hold for this choice of p_1, p_2, p, where in* (iv) *we replace the requirement $r > 2$ with $r > 2 + \varepsilon$.*

It may be possible to refine the range of p_1, p_2 here to match that in Corollary 11.3 or Lemma 11.1 by a more careful argument, but we will not attempt to do so here.

Proof (*sketch*). We repeat the proof of Theorem 1.17. By the arguments in Section 3, it suffices to show that Theorem 3.9 holds for the indicated choice of p_1, p_2, p. We then repeat the reductions in Section 5 that were used to reduce Theorem 3.9 to Theorem 5.30. The only differences are that (1.6) is replaced by the more general Lemma 11.1 (which, in particular, is applicable for $(\frac{1}{p_1}, \frac{1}{p_2})$ sufficiently close to $(\frac{1}{2}, \frac{1}{2})$) and uses the quasi–triangle inequality in place of the triangle inequality when $p < 1$ (adjusting the exponent 10 appearing in the

argument if necessary). It then suffices to establish Theorem (5.30) for $(\frac{1}{p_1}, \frac{1}{p_2})$ in a neighborhood of $(\frac{1}{2}, \frac{1}{2})$. In fact, it suffice to establish the cruder estimate

$$\|(\tilde{A}_N(F_N, G_N))_{N\in\mathbb{I}}\|_{\ell^p(\mathbb{Z};\ell^\infty)} \lesssim_{C_3} 2^{O(\max(l,s_1,s_2))}\|f\|_{\ell^{p_1}(\mathbb{Z})}\|g\|_{\ell^{p_2}(\mathbb{Z})}$$

for $(\frac{1}{p_1}, \frac{1}{p_2})$ in a neighbourhood of $(\frac{1}{2}, \frac{1}{2})$, since the claim then follows by interpolation with the $p_1 = p_2 = 2$ case of Theorem 5.30 and reducing the size of the neighborhood in an ε-dependent fashion. (Here we use the interpolation theory[10] of variational norms, as well as the equivalence $\boldsymbol{V}^\infty \equiv \ell^\infty$.)

The contribution of the small scales $\mathbb{I}_\leq$ can now be crudely handled by Lemma 11.1 and the quasi-triangle inequality (3.7) (since we are now willing to concede factors of $2^{O(l)}$). Hence we may work entirely with large scales $\mathbb{I}_>$. It is not difficult to verify that Proposition 7.13 extends to the non-Banach regime $p < 1$ (basically because Lemma 4.20 does, and because one can freely lose powers of q in that proposition). Applying the arguments in Section 7 with suitable changes, we reduce to showing that the $\ell^p(\mathbb{Z};\ell^\infty)$ norm of (7.25) is bounded by

$$\lesssim_{C_3} 2^{O(\max(l,s_1,s_2))}\|f\|_{\ell^{p_1}(\mathbb{Z})}\|g\|_{\ell^{p_2}(\mathbb{Z})}$$

for $(\frac{1}{p_1}, \frac{1}{p_2})$ in a neighborhood of $(\frac{1}{2}, \frac{1}{2})$.

In the non-Banach regime we are no longer able to remove the integration in t; instead we crudely replace it by a supremum norm. In lieu of Theorem 7.28, it will now suffice to show that

$$\|H\|_{L^p(\mathbb{A}_\mathbb{Z};\boldsymbol{V}^r)} \lesssim_{C_3} 2^{O(\max(l,s_1,s_2))}\|F_\mathbb{A}\|_{L^{p_1}(\mathbb{A}_\mathbb{Z})}\|G_\mathbb{A}\|_{L^{p_2}(\mathbb{A}_\mathbb{Z})},$$

where H is the maximal operator

$$H := \sup_{t\in[1/2,1]} |\mathrm{B}_{1\otimes m_{\hat{\mathbb{Z}}}}(\mathrm{T}_{\varphi_{N,t}\otimes 1}F_\mathbb{A}, \mathrm{T}_{\tilde{\varphi}_{N,t}\otimes 1}G_\mathbb{A})|.$$

(Here we implicitly use the fact that Theorem 4.18 continues to hold in the range $p < 1$.) By a variant of (9.2), each slice H_x of H at some $x \in \mathbb{R}$ is given by

$$H_x = \sup_{t\in[1/2,1]} |A_{\hat{\mathbb{Z}}}((T_{\varphi_{N,t}\otimes 1}F_\mathbb{A})_x, (T_{\tilde{\varphi}_{N,t}\otimes 1}G_\mathbb{A})_x)|.$$

We crudely bound

$$T_{\varphi_{N,t}\otimes 1}F_\mathbb{A} \lesssim_{C_3} 2^{\max(0,s_1)}\mathrm{M_{HL}}F_\mathbb{A},$$

$$T_{\tilde{\varphi}_{N,t}\otimes 1}G_\mathbb{A} \lesssim_{C_3} 2^{\max(0,s_2)}\mathrm{M_{HL}}G_\mathbb{A},$$

where $\mathrm{M_{HL}}$ denotes the Hardy–Littlewood maximal operator in the $\mathbb{R}$ variable, so that

$$H_x \lesssim_{C_3} 2^{O(\max(l,s_1,s_2))}A_{\hat{\mathbb{Z}}}((\mathrm{M_{HL}}F_\mathbb{A})_x, (\mathrm{M_{HL}}G_\mathbb{A})_x).$$

[10] See, for instance, [69] for an overview of this interpolation theory.

From the Hardy–Littlewood inequality and the Fubini–Tonelli theorem, it now suffices to establish the estimate

$$\|A_{\hat{\mathbb{Z}}}(F,G)\|_{L^p(\mathbb{A}_\mathbb{Z})} \lesssim \|F\|_{L^{p_1}(\mathbb{A}_\mathbb{Z})}\|G\|_{L^{p_2}(\mathbb{A}_\mathbb{Z})}$$

for any $F \in L^{p_1}(\mathbb{A}_\mathbb{Z})$, $G \in L^{p_2}(\mathbb{A}_\mathbb{Z})$. When $p \geq 1$, this follows from Hölder's inequality and the triangle inequality. For $p < 1$, we can interpolate the $p \geq 1$ estimate with (10.2) and conclude that

$$\|A_{\hat{\mathbb{Z}}}(F,G)\|_{L^1(\mathbb{A}_\mathbb{Z})} \lesssim \|F\|_{L^{p_1}(\mathbb{A}_\mathbb{Z})}\|G\|_{L^{p_2}(\mathbb{A}_\mathbb{Z})}$$

for all $(\frac{1}{p_1}, \frac{1}{p_2})$ sufficiently close to $(\frac{1}{2}, \frac{1}{2})$. The claim now follows from Hölder's inequality. □

12. **Unboundedness of quadratic variation**

In this section we show that the quadratic variation of polynomial averages is unbounded in any Lebesgue space norm. The counterexample already applies in the linear setting:

PROPOSITION 12.1 (Unboundedness of V^2). *Let $P(\mathrm{n}) \in \mathbb{Z}[\mathrm{n}]$ be a non-constant polynomial, and let $0 < p \leq \infty$. Let $\mathbb{I} \subseteq \mathbb{Z}_+$ be an infinite set. Then for every $C > 0$, there exists a measure-preserving system (X, μ, T) of total measure 1 and $f \in L^\infty(X)$ with $\|f\|_{L^\infty(X)} \leq 1$ such that*

$$\|(A^{P(\mathrm{n})}_{N,X}(f))_{N\in\mathbb{I}}\|_{L^p(X;V^2)} > C.$$

We remark that the case $p = 2$ of this proposition (with f controlled in L^2 rather than L^∞) was established by the first author in [54]. (The argument there is given for $P(\mathrm{n}) = \mathrm{n}^2$, but extends easily to more general polynomials.) This result relied on a previous result of Lewko and Lewko [59], who in turn invoked a result of Jones and Wang [51]. It turns out that by appealing to the latter results directly, we can handle all values of p, answering [54, Conj. 1] in the affirmative.

Proof. Suppose for contradiction that this were not the case. Then we would have the variational inequality

$$\|(A^{P(\mathrm{n})}_{N,X}(f))_{N\in\mathbb{I}}\|_{L^p(X;V^2)} \leq C\|f\|_{L^\infty(X)} \tag{12.2}$$

for every measure-preserving system (X, μ, T) and every $f \in L^\infty(X)$.

We apply this inequality to the following multidimensional system in which the different components of the shift have radically different mixing times (so that the averages $A^{P(\mathrm{n})}_{N,X}$ behave like martingale expectation operators). Set $X = \mathbb{T}^K$ for some $K \in \mathbb{Z}_+$ with Haar probability measure μ, and let $f\colon X \to \mathbb{C}$ be a smooth function. Fix a sequence $\alpha_1, \alpha_2, \ldots, \alpha_K$ of real numbers that are linearly independent over $\mathbb{Q}$; e.g., one could take $\alpha_i := \log p_i$ where p_i is the

i^{th} prime. Let $N_1 < \cdots < N_K$ be distinct elements of $\mathbb{I}$, and consider the shift map

$$T(x_1, \dots, x_K) \coloneqq \left(x_1 + \frac{\alpha_1}{N_1^{d+1}}, \dots, x_K + \frac{\alpha_K}{N_K^{d+1}} \right),$$

where d is the degree of P. Then for any $k \in [K]$, we have

$$A^{P(\mathrm{n})}_{N_k, \mathbb{T}^K} f(x_1, \dots, x_K) = \mathbb{E}_{n \in [N_k]} f \left(x_1 + \frac{\alpha_1 P(n)}{N_1^{d+1}}, \dots, x_K + \frac{\alpha_K P(n)}{N_K^{d+1}} \right).$$

Let $\varepsilon > 0$. If we assume for each $k \in [K]$ that N_k is sufficiently large depending on $\varepsilon, N_1, \dots, N_{k-1}, d, P, f$, then we have

$$\begin{aligned} &f \left(x_1 + \frac{\alpha_1 P(n)}{N_1^{d+1}}, \dots, x_K + \frac{\alpha_K P(n)}{N_K^{d+1}} \right) \\ &\quad = f \left(x_1 + \frac{\alpha_1 P(n)}{N_1^{d+1}}, \dots, x_{k-1} + \frac{\alpha_{k-1} P(n)}{N_{k-1}^{d+1}}, x_k, \dots, x_K \right) + O(\varepsilon) \end{aligned}$$

for all $n \in [N_k]$ and $(x_1, \dots, x_K) \in X$, and thus

$$\begin{aligned} &A^{P(\mathrm{n})}_{N_k, \mathbb{T}^K} f(x_1, \dots, x_K) \\ &\quad = \mathbb{E}_{n \in [N_k]} f \left(x_1 + \frac{\alpha_1 P(n)}{N_1^{d+1}}, \dots, x_{k-1} + \frac{\alpha_{k-1} P(n)}{N_{k-1}^{d+1}}, x_k, \dots, x_K \right) + O(\varepsilon). \end{aligned}$$

Because $\alpha_1, \dots, \alpha_{k-1}$ are linearly independent, a standard application of the Weyl equidistribution theorem shows that the sequence

$$\left(\frac{\alpha_1 P(n)}{N_1^{d+1}} \bmod 1, \dots, \frac{\alpha_{k-1} P(n)}{N_{k-1}^{d+1}} \bmod 1 \right)$$

is equidistributed over the torus $\mathbb{T}^{k-1}$. Thus, if N_k is chosen large enough, we have

$$A^{P(\mathrm{n})}_{N_k, \mathbb{T}^K} f(x) = \mathbb{E}_k f(x) + O(\varepsilon)$$

for all $k \in [K]$ and $x \in X$, where $\mathbb{E}_k f$ is the conditional expectation

$$\mathbb{E}_k f(x_1, \dots, x_K) \coloneqq \int_{\mathbb{T}^{k-1}} f(y_1, \dots, y_{k-1}, x_k, \dots, x_K)\, dy_1 \dots dy_{k-1}.$$

Taking variations, we conclude that

$$\| (A^{P(\mathrm{n})}_{N_k, \mathbb{T}^K} - \mathbb{E}_k f)_{k \in [K]} \|_{L^p(\mathbb{T}^K; V^2)} \lesssim_K \varepsilon,$$

which from (12.2) and the triangle inequality (or quasi–triangle inequality (3.7)) gives

$$\| (\mathbb{E}_k f)_{k \in [K]} \|_{L^p(\mathbb{T}^K; V^2)} \lesssim_{C,p} \| f \|_{L^\infty(\mathbb{T}^K)} + O_K(\varepsilon).$$

Sending $\varepsilon \to 0$ (noting that the left-hand side does not depend on ε-dependent quantities such as $N_1, \dots, N_K$), we conclude that

$$\|(\mathbb{E}_k f)_{k\in[K]}\|_{L^p(\mathbb{T}^K;V^2)} \lesssim_{C,p} \|f\|_{L^\infty(\mathbb{T}^K)}$$

for any smooth $f \in L^\infty(\mathbb{T}^K)$. Taking limits, we see that we can drop the hypothesis that f is smooth.

We now define a map $\pi\colon \mathbb{T}^K \to [0,1)$ by the formula

$$\pi(x_1 \bmod 1, \dots, x_K \bmod 1) := \sum_{k\in[K]} \frac{\lfloor 2x_k \rfloor}{2^{K-k+1}}$$

for $x_1, \dots, x_K \in [0,1)$. It is not difficult to see that π pushes forward Haar measure on $\mathbb{T}^K$ to Lebesgue measure on $[0,1)$. Furthermore, if $\tilde{f} \in L^\infty([0,1))$, then

$$\mathbb{E}_k(\tilde{f} \circ \pi) = (\tilde{\mathbb{E}}_k \tilde{f}) \circ \pi$$

almost everywhere on $\mathbb{T}^K$, where $\tilde{\mathbb{E}}_k$ are the martingale projections

$$\tilde{\mathbb{E}}_k \tilde{f}(x) := 2^k \int_{(j-1)/2^k}^{j/2^k} f(y)\, dy$$

whenever $j \in [2^k]$ and $x \in [(j-1)/2^k, j/2^k)$. From this we conclude that

$$\|(\tilde{\mathbb{E}}_k \tilde{f})_{k\in[K]}\|_{L^p([0,1);V^2)} \lesssim_{C,p} \|\tilde{f}\|_{L^\infty([0,1))}$$

for all $K \in \mathbb{N}$ and $f \in L^\infty([0,1))$. Taking $K \to \infty$ and using monotone convergence, we conclude that

$$\|(\tilde{\mathbb{E}}_k \tilde{f})_{k\in\mathbb{N}}\|_{L^p([0,1);V^2)} \lesssim_{C,p} \|\tilde{f}\|_{L^\infty([0,1))}.$$

But this contradicts [51, Prop. 8.1]. □

Remark 12.3. By considering a suitable product system, one can then construct a single measure-preserving system (X, μ, T) of total measure 1 such that the vector-valued operator $f \mapsto (A_N^{P(\mathrm{n})}(f))_{N\in\mathbb{I}}$ is unbounded from $L^2(X)$ to $L^p(X; V^2)$. It is likely that one can sharpen the construction further to find a single $f \in L^2(X)$ for which $\|(A_N^{P(\mathrm{n})}(f))_{N\in\mathbb{I}}\|_{V^2} = +\infty$ almost everywhere, but we will not do so here.

By setting all but one function equal to the constant function 1, and using the monotonicity of variational norms and L^p norms, we obtain

COROLLARY 12.4 (Failure of variational estimate for $r \le 2$). *Let $P_1, \dots, P_k \in \mathbb{Z}[\mathrm{n}]$ be polynomials, not all constant, and let $0 < p_1, \dots, p_k, p \le \infty$ and $0 < r \le 2$. Let $\mathbb{I} \subseteq \mathbb{Z}_+$ be an infinite set. Then there does not exist any constant $C > 0$ for which one has the estimate*

$$\|(A_{N,X}^{P_1(\mathrm{n}),\dots,P_k(\mathrm{n})}(f_1, \dots, f_k))_{N\in I}\|_{L^p(X;V^r)} \le C\|f_1\|_{L^{p_1}(X)} \cdots \|f_k\|_{L^{p_k}(X)}$$

for all measure-preserving systems $X = (X, \mu, T)$ *of total mass one, and all* $f_1 \in L^{p_1}(X), \dots, f_k \in L^{p_k}(X)$.

Applying Proposition 3.2 in the contrapositive, we see that we similarly obtain a counterexample for the integer shift system in the Hölder exponent case $\frac{1}{p_1} + \cdots + \frac{1}{p_k} = \frac{1}{p}$, and we can replace A_N by $\tilde{A}_N$ in the Banach exponent case $p_1, \dots, p_k \geq 1$.

Appendix A. Ionescu–Wainger theory

In this appendix we review some number-theoretic and Fourier-analytic constructions of Ionescu and Wainger [47] that allow one to apply Fourier projections to "major arcs" with good multiplier estimates. See also [65], [68] for further development of the Ionescu–Wainger theory, and see [76] for a recent discussion of the role of superorthogonality in that theory. We will loosely follow the presentation in [68]. A new notational innovation is the introduction of the notion of the height $\mathrm{h}(\alpha)$ of a profinite frequency $\alpha \in \mathbb{Q}/\mathbb{Z}$.

Throughout this appendix we fix a small quantity $\rho > 0$. (In the main paper it is set by the formula (5.1).) Let C^0_ρ be a sufficiently large quantity depending on ρ. If $l \leq C^0_\rho$, we define

$$P_{\leq l} := [2^l].$$

For $l > C^0_\rho$, we define $P_{\leq l}$ differently. We first define the natural number

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

and for any natural number $l \in \mathbb{N}$, we set

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

Then for $l > C^0_\rho$, we define the set

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

where

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

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

We observe that (for C^0_ρ large enough) one has

$$[2^l] \subset P_{\leq l} \tag{A.1}$$

for all l. This is trivial for $l \leq C^0_\rho$. Now suppose that $l > C^0_\rho$ and $q \in [2^l]$. Observe that there are at most D primes larger than N_0 that can divide q, and each such prime can divide q at most D times, so the product of all these

primes (with multiplicity) lies in $W_{\leq l} \cup \{1\}$. By the fundamental theorem of arithmetic, the claim will now follow if one can show that $p^j | Q_0$ whenever $p \leq N_0$ and $p^j | q$. Since $j \leq \frac{\log q}{\log p} \leq \frac{l}{\log p}$ (recall our convention that log is to base 2), and p divides $N_0!$ at least $\lfloor \frac{N_0}{p} \rfloor$ times, it suffices to establish the inequality

$$\frac{l}{\log p} \leq D \left\lfloor \frac{N_0}{p} \right\rfloor.$$

Since $D > \frac{2}{\rho} \geq \frac{l}{\log N_0}$, it suffices to show that

$$\frac{\log N_0}{\log p} \leq (1 + \varepsilon_\rho) \left\lfloor \frac{N_0}{p} \right\rfloor$$

for $2 \leq p \leq N_0$, where ε_ρ is the positive quantity $\varepsilon_\rho := \frac{\rho D}{2} - 1$. If we set $n := \lfloor \frac{N_0}{p} \rfloor$, then $n \in [N_0/2]$ and $\frac{\log N_0}{\log p} \leq \frac{\log N_0}{\log N_0 - \log(n+1)}$, so after some rearranging we reduce to showing that

$$\log(n+1) \leq \left(1 - \frac{1}{(1+\varepsilon_\rho)n}\right) \log N_0$$

for all $n \in [N_0/2]$. But this can be easily checked if C_ρ^0 (and hence N_0) is sufficiently large depending on ρ. (One can, for instance, check the cases $1 \leq n \leq N_0^{1/2}$ and $N_0^{1/2} < n \leq N_0/2$ separately.)

We now see that the $P_{\leq l}$ are non-decreasing in l with $\bigcup_{l \in \mathbb{N}} P_{\leq l} = \mathbb{Z}_+$. We can therefore define the Ionescu–Wainger height $\mathrm{h}(\alpha) = \mathrm{h}_\rho(\alpha)$ of an arithmetic frequency $\frac{a}{q} \bmod 1$, with $q \in \mathbb{Z}_+$ and $a \in [q]^\times$, by the formula

$$\mathrm{h}\left(\frac{a}{q} \bmod 1\right) := \inf\{2^l : l \in \mathbb{N}, q \in P_{\leq l}\}.$$

Now we prove Lemma 5.2. The claim (i) is immediate from (A.1), with the final claim concerning $\frac{1}{p}\mathbb{Z}/\mathbb{Z}$ following from direct inspection of definitions. For the first part of (ii), we observe that

$$(\mathbb{Q}/\mathbb{Z})_{\leq l} = \bigcup_{q \in P_{\leq l}} \frac{1}{q}\mathbb{Z}/\mathbb{Z}, \tag{A.2}$$

so it suffices to show that $q \lesssim_\rho 2^{2^{\rho l}}$ for all $q \in P_{\leq l}$. For $l > C_\rho^0$, we have from definition that

$$q \leq Q_0 (2^l)^{D^2} \leq N_0^{DN_0} 2^{D^2 l} \lesssim_\rho 2^{\rho^{-1} 2^{\rho l/2} + \rho^{-2} l}$$

giving the claim; in fact, we obtain the slightly sharper bound

$$q \lesssim_\rho 2^{O_\rho(2^{\rho l/2})}. \tag{A.3}$$

For the second claim, we need to show that

$$Q_{\leq l} := \mathrm{lcm}(q \in \mathbb{Z}_+ : q \in P_{\leq l}) \lesssim_\rho 2^{O(2^l)}. \tag{A.4}$$

The claim is trivial for $l \leq C_\rho^0$. For $l > C_\rho^0$, we have

$$\operatorname{lcm}(q \in \mathbb{Z}_+ : q \in P_{\leq l}) = Q_0 \prod_{p \in (N_0^{(l)}, 2^l] \cap \mathbb{P}} p^D.$$

From Mertens' theorem we have

$$\prod_{p \in (N_0^{(l)}, 2^l] \cap \mathbb{P}} p \lesssim 2^{O(2^l)}$$

and

$$Q_0 \leq N_0^{DN_0} \lesssim 2^{O_\rho(2^{\rho l})},$$

giving the claim. Claim (iii) follows from (A.2) and (A.3). This proves Lemma 5.2.

To establish Theorem 5.7, we observe from Lemma 5.2(ii) and (A.3) that the elements of $(\mathbb{Q}/\mathbb{Z})_{\leq l}$ are separated from each other by $\gtrsim_\rho 2^{-O_\rho(2^{\rho l/2})}$, giving the non-aliasing claim. The claim (5.9) follows[11] from [68, Th. 2.1] (specialized to the one-dimensional case); various special cases of this theorem were previously established in [47]; see also Remark 5.10(i). Note that on the right-hand side one can use the scalar norm rather than the vector-valued norm thanks to the Marcinkiewicz–Zygmund inequality (or Khintchine's inequality). Finally, the claim for the multipliers (5.5) follows from (5.9) and the triangle inequality.

Now we prove Lemma 5.17. The Fourier support properties are clear from inspection and the disjointness of the individual major arcs. The contraction property on ℓ^2 follows from Plancherel's theorem because the symbol $\mathcal{P}\eta_{\leq k}$ is bounded pointwise by 1. To obtain the bound (5.18), by interpolation we may assume that q is either an even integer or the dual of an even integer. Then it suffices from Theorem 5.7 to establish the bound

$$\|\mathrm{T}_{\eta_{\leq k}}\|_{L^q(\mathbb{R}) \to L^q(\mathbb{R})} \lesssim_q 1. \tag{A.5}$$

But this follows from Lemma 4.20 (with $r = 2^k$).

Finally we establish (5.19). It suffices to establish the bound

$$\|\Pi_{\leq l, \leq k} f\|_{\ell^q(\{n \in \mathbb{Z} : \operatorname{dist}(n, I) > 2^{m-k}\})} \lesssim_M 2^{-Mm} \|f\|_{\ell^q(I)}$$

for any $m \in \mathbb{Z}_+$. By interpolation we may assume q is either an even integer or the dual of an even integer. By adjusting constants in the definition (5.8) of good major arcs if necessary, we may assume that

$$k \leq -2v,$$

where

$$v := \lfloor C_\rho 2^{\rho l} \rfloor.$$

[11] The factor $\langle l \rangle$ in this theorem was recently removed in [80].

We split

$$\eta_{\le k} := \eta^{(1)}_{\le k} + \eta^{(2)}_{\le k} + \eta^{(3)}_{\le k},$$

where $\eta^{(1)}_{\le k}, \eta^{(2)}_{\le k}, \eta^{(3)}_{\le k} \in \boldsymbol{S}(\mathbb{R})$ are the functions

$$\begin{aligned}
\eta^{(1)}_{\le k} &:= \mathcal{F}_{\mathbb{R}}(\eta_{\le m-k}\mathcal{F}_{\mathbb{R}}^{-1}\eta_{\le k}),\\
\eta^{(2)}_{\le k} &:= \mathcal{F}_{\mathbb{R}}((1-\eta_{\le m-k})\mathcal{F}_{\mathbb{R}}^{-1}\eta_{\le k})\eta_{\le -v},\\
\eta^{(3)}_{\le k} &:= \mathcal{F}_{\mathbb{R}}((1-\eta_{\le m-k})\mathcal{F}_{\mathbb{R}}^{-1}\eta_{\le k})(1-\eta_{\le -v}),\\
&= -\mathcal{F}_{\mathbb{R}}(\eta_{\le m-k}\mathcal{F}_{\mathbb{R}}^{-1}\eta_{\le k})(1-\eta_{\le -v}).
\end{aligned}$$

We can then decompose

$$\Pi_{\le l,\le k} f = \mathrm{T}^{\le l}_{\eta_{\le k}} f = \mathrm{T}^{\le l}_{\eta^{(1)}_{\le k}} f + \mathrm{T}^{\le l}_{\eta^{(2)}_{\le k}} f + \mathrm{T}^{\le l}_{\eta^{(3)}_{\le k}} f.$$

Note that the inverse Fourier transform of $\eta^{(1)}_{\le k}$ is supported in $[-2^{m-k}, 2^{m-k}]$, and hence $\mathrm{T}^{\le l}_{\eta^{(1)}_{\le k}} f$ vanishes on the region $\{n \in \mathbb{Z} : \mathrm{dist}(n, I) > 2^{m-k}\}$. For $\eta^{(2)}_{\le k}$, we use Theorem 5.7 (and the fact that $(k, -v)$ has good major arcs), (A.5), Young's inequality, and a rescaling to bound

$$\begin{aligned}
\|\mathrm{T}^{\le l}_{\eta^{(2)}_{\le k}} f\|_{\ell^q(\mathbb{Z})} &\lesssim_q \langle l\rangle \|\mathrm{T}_{\eta^{(2)}_{\le k}}\|_{L^q(\mathbb{R})\to L^q(\mathbb{R})}\|f\|_{\ell^q(I)}\\
&\lesssim_q \langle l\rangle \|\mathrm{T}_{\mathcal{F}_{\mathbb{R}}((1-\eta_{\le m-k})\mathcal{F}_{\mathbb{R}}^{-1}\eta_{\le k})}\|_{L^q(\mathbb{R})\to L^q(\mathbb{R})}\|f\|_{\ell^q(I)}\\
&\lesssim_q \langle l\rangle \|(1-\eta_{\le m-k})\mathcal{F}_{\mathbb{R}}^{-1}\eta_{\le k}\|_{L^1(\mathbb{R})}\|f\|_{\ell^q(I)}\\
&\lesssim_q \langle l\rangle \|\mathcal{F}_{\mathbb{R}}^{-1}\eta\|_{L^1(\mathbb{R}\setminus[-2^{m-1},2^{m-1}])}\|f\|_{\ell^q(I)}.
\end{aligned}$$

Hence the contribution of this term is acceptable by the rapid decrease of $\mathcal{F}_{\mathbb{R}}^{-1}\eta$.

Finally, for $\eta^{(3)}_{\le k}$, we use Lemma 4.20(i) (with $r = 2^k$) and Lemma 5.2(iii) to bound

$$\Big\|\mathrm{T}^{\le l}_{\eta^{(3)}_{\le k}} f\Big\|_{\ell^q(\mathbb{Z})} \lesssim_{C_1,q} 2^{O(2^{\rho l})}\|f\|_{\ell^q(I)} \sup_{0\le j\le 2}\int_{\mathbb{R}} 2^{k(1-j)}\Big|\frac{d^j}{d\xi^j}\eta^{(3)}_{\le k}(\xi)\Big|\,d\xi.$$

Direct calculation using the rapid decay of $\mathcal{F}_{\mathbb{R}}\eta$ shows that

$$\int_{\mathbb{R}} 2^{k(1-j)}\Big|\frac{d^j}{d\xi^j}\eta^{(3)}_{\le k}(\xi)\Big|\,d\xi \lesssim_M 2^{M(k-m+v)} \lesssim 2^{-Mm}2^{-MC_\rho 2^{\rho l}},$$

and hence the contribution of this term is also acceptable (taking C_ρ large enough). This concludes the proof of Lemma 5.17.

Appendix B. Shifted Calderón–Zygmund theory

In this appendix we review some standard shifted Calderón–Zygmund estimates, of the sort that appear for instance in [63, Lemma 4.8, pp. 346]. For our applications, we will need a vector-valued version of these estimates.

THEOREM B.1 (Shifted Calderón–Zygmund estimates). *Let $\mathbb{D}$ be a finite λ-lacunary set for some $\lambda > 1$, and let $A > 0$, $C > 0$, $d \geq 1$, and $K \geq 1$. For each $N \in \mathbb{D}$, let $\varphi_N \in \boldsymbol{S}(\mathbb{R})$ be a function of the form*

$$\varphi_N(\xi) := \psi(AN^d\xi)e(\lambda_N AN^d\xi)$$

for some $\lambda_N \in [-2^K, 2^K]$, where $\psi \in \boldsymbol{S}(\mathbb{R})$ vanishes at the origin and is supported on $[-C, C]$ for some $C > 0$, obeying the derivative estimates

$$\left|\frac{d^j}{d\xi^j}\psi(\xi)\right| \leq C$$

for all $j = 0, 1, 2$ and $\xi \in \mathbb{R}$. Then for any $1 < p < \infty$ and any separable Hilbert space $(H, \|\cdot\|_H)$, one has

$$\|\mathrm{T}_{\sum_{N\in\mathbb{D}} \epsilon_N \varphi_N}\|_{L^p(\mathbb{R};H)\to L^p(\mathbb{R};H)} \lesssim_{C,\lambda,d,p} K$$

for any complex numbers $\epsilon_N, N \in \mathbb{D}$ with $|\epsilon_N| \leq 1$; in particular, by Khintchine's inequality,

$$\|(\mathrm{T}_{\varphi_N})_{N\in\mathbb{D}}\|_{L^p(\mathbb{R};H)\to L^p(\mathbb{R};\ell^2(\mathbb{D};H))} \lesssim_{C,\lambda,d,p} K.$$

Proof (sketch). Let $\varphi := \sum_{N\in\mathbb{D}} \epsilon_N \varphi_N$. From the hypotheses on ψ, one has the bound

$$|\psi(\xi)| \lesssim_C |\xi| \mathbb{1}_{|\xi|\leq C},$$

and hence from the triangle inequality, one has $\|\varphi\|_{L^\infty(\mathbb{R})} \lesssim_{C,\lambda} 1$. The $p = 2$ case of the theorem then follows from Plancherel's theorem. By duality it then suffices to establish the $1 < p < 2$ case, and by Marcinkiewicz interpolation, it suffices to prove the weak-type $(1, 1)$ bound

$$|\{x \in \mathbb{R} : \|\mathrm{T}_\varphi f(x)\|_H \geq \alpha\}| \lesssim_{C,\lambda,d} \frac{K}{\alpha}\|f\|_{L^1(\mathbb{R};H)}$$

for $f \in L^1(\mathbb{R}; H)$ and $\alpha > 0$. We perform a vector-valued Calderón–Zygmund decomposition $f = g + \sum_{I\in\mathcal{D}} b_I$, where $\|g\|^2_{L^2(\mathbb{R};H)} \lesssim \|f\|_{L^1(\mathbb{R};H)}\alpha$, I ranges over a collection of dyadic intervals $\mathcal{D}$ with

$$\sum_{I\in\mathcal{D}} |I| \lesssim_{C,\lambda} \frac{1}{\alpha}\|f\|_{L^1(\mathbb{R};H)},$$

and $b_I \in L^1(\mathbb{R}; H)$ is supported on I with mean zero and

$$\|b_I\|_{L^1(\mathbb{R};H)} \lesssim |I|. \tag{B.2}$$

By the previous inequality it suffices to prove

$$\left|\left\{x \in \Big(\bigcup_{I\in\mathcal{D}} 100I\Big)^c : \|\mathrm{T}_\varphi f(x)\|_H \geq \alpha\right\}\right| \lesssim_{C,\lambda,d} \frac{K}{\alpha}\|f\|_{L^1(\mathbb{R};H)},$$

where aI is the interval centered at I of $a>0$ times the length. By the triangle inequality and Markov's inequality, it thus suffices to show that

$$\int_{(100I)^c} \|\mathrm{T}_\varphi b_I(x)\|_H \, dx \lesssim_{C,\lambda} K|I| \tag{B.3}$$

for each $I \in \mathcal{D}$. We may expand

$$\mathrm{T}_\varphi b_I(x) = \sum_{N\in\mathbb{D}} \epsilon_N \int_{\mathbb{R}} (AN^d)^{-1} \mathcal{F}_{\mathbb{R}}^{-1}\psi\Big(\frac{x-y-\lambda_N AN^d}{AN^d}\Big) b_I(y)\, dy.$$

We may assume that $I \in \mathcal{D}$ is centered at the origin, and exploiting the fact that b_I has mean zero, we may dominate the left-hand side of (B.3) by

$$\sum_{N\in\mathbb{D}} \int_{(100I)^c} \int_I \frac{1}{AN^d} \Big| \mathcal{F}_{\mathbb{R}}^{-1}\psi\Big(\frac{x-\lambda_N AN^d - y}{AN^d}\Big) - \mathcal{F}_{\mathbb{R}}^{-1}\psi\Big(\frac{x-\lambda_N AN^d}{AN^d}\Big)\Big| \|b_I(y)\|_H \, dydx.$$

So by (B.2) it suffices to show that

$$\sum_{N\in\mathbb{D}} \int_{(100I)^c} \frac{1}{AN^d} \Big| \mathcal{F}_{\mathbb{R}}^{-1}\psi\Big(\frac{x-\lambda_N AN^d - y}{AN^d}\Big) - \mathcal{F}_{\mathbb{R}}^{-1}\psi\Big(\frac{x-\lambda_N AN^d}{AN^d}\Big)\Big| dx \lesssim_{C,\lambda} K$$

for all $y \in I$.

Fix y, I. We perform a partition

$$\mathbb{D} = \mathbb{D}_{\mathrm{low}} \cup \mathbb{D}_{\mathrm{medium}} \cup \mathbb{D}_{\mathrm{high}},$$

where $\mathbb{D}_{\mathrm{low}}$ consists of those spatial scales $N \in \mathbb{D}$ that are "low frequency" (or "coarse scale") in the sense that $|I| \leq AN^d$, $\mathbb{D}_{\mathrm{medium}}$ consists of those spatial scales $N \in \mathbb{D}$ that are "medium frequency" (or "medium scale") in the sense that $\lambda_N^{-1}|I| \leq AN^d < |I|$, and $\mathbb{D}_{\mathrm{high}}$ consists of those spatial scales $N \in \mathbb{D}$ that are "high frequency" (or "fine scale") in the sense that $AN^d < \lambda_N^{-1}|I|$.

The expression

$$\mathcal{F}_{\mathbb{R}}^{-1}\psi\Big(\frac{x-\lambda_N AN^d - y}{AN^d}\Big) - \mathcal{F}_{\mathbb{R}}^{-1}\psi\Big(\frac{x-\lambda_N AN^d}{AN^d}\Big)$$

can be bounded by $O_C(\langle \frac{x}{AN^d}\rangle^{-2})$ in the high-frequency case $N \in \mathbb{D}_{\mathrm{high}}$ from the triangle inequality and the hypotheses $y \in I$, $x \in (100I)^c$, by $O_C(\langle \frac{x}{AN^d} - \lambda_N\rangle^{-2})$ in the medium-frequency case $N \in \mathbb{D}_{\mathrm{medium}}$ from the triangle inequality alone, and by $O_C(\frac{|I|}{AN^d}\langle \frac{x}{AN^d} - \lambda_N\rangle^{-2})$ in the low-frequency case using the

mean-value theorem. The claim then follows from direct computation and the hypothesis $|\lambda_N| \leq 2^K$. □

Appendix C. Concentration estimates on polynomials

In this appendix we work in a p-adic field $\mathbb{Q}_p = \bigcup_{n\in\mathbb{N}} p^{-j}\mathbb{Z}_p$ for $p \in \mathbb{P}$, although much of the discussion here would also extend with minor changes to the real numbers $\mathbb{R}$ or (after adjusting some exponents by factors of two) the complex numbers $\mathbb{C}$, and the reader may wish to work with the real case first to build intuition. We have a norm on the p-adics defined by $|x| := p^{-\nu_p(x)}$, where ν_p is the usual p-valuation (with the usual convention $|0| = 0$), as well as a Haar measure $\mu_{\mathbb{Q}_p}$ on $\mathbb{Q}_p$ with the following properties for any $x, y \in \mathbb{Q}_p$ and $r \in p^{\mathbb{Z}} := \{p^n : n \in \mathbb{Z}\}$:

(i) (ultratriangle inequality). $|x + y| \leq \max(|x|, |y|)$;
(ii) (multiplicativity). $|xy| = |x||y|$;
(iii) (nondegeneracy). $|x| \geq 0$, with equality if and only if $x = 0$;
(iv) (dimension one). $\mu_{\mathbb{Q}_p}(B(x, r)) = r$, where $B(x, r) := \{y \in \mathbb{Q}_p : |y - x| \leq r\}$ is the usual ball.

Note that if P is a polynomial with coefficients in $\mathbb{Q}_p$, thus

$$P(x) = a_d x^d + \cdots + a_1 x + a_0$$

for some $a_d, \ldots, a_0 \in \mathbb{Q}_p$, then one can define the derivative P' algebraically by the usual formula

$$P'(x) := d a_d x^{d-1} + \cdots + 2a_2 x + a_1.$$

We then have the following basic estimates on the distribution of p-adic polynomials.

Proposition C.1 (Distribution of p-adic polynomials). *Let $P(x) = a_d x^d + \cdots + a_0$ be a polynomial of degree $d \geq 1$ with coefficients in $\mathbb{Q}_p$. Let $r \in p^{\mathbb{Z}}$, and let Ω be the level set*

$$\Omega := \{x \in \mathbb{Q}_p : |P(x)| \leq r\}.$$

(i) (Bernstein inequality). *One can cover Ω by $O_d(1)$ balls B, such that on each ball B one has*

$$\sup_{x\in B} |P'(x)| \lesssim_d \frac{r}{\mu_{\mathbb{Q}_p}(B)}.$$

(ii) (Van der Corput estimate). *We have*

$$\mu_{\mathbb{Q}_p}(\Omega) \lesssim_d \left(\frac{r}{|a_d|}\right)^{1/d}.$$

In fact Ω is covered by $O_d(1)$ balls of radius $\left(\frac{r}{|a_d|}\right)^{1/d}$.

(iii) (Distributional estimate). *If $d \geq 2$, and if $f\colon \mathbb{Q}_p \to [0,+\infty)$ is the function*

$$f(y) \coloneqq \frac{1}{r}\mu_{\mathbb{Q}_p}(\{x \in \mathbb{Q}_p : |P(x)-y| \leq r\}),$$

then

$$\mu_{\mathbb{Q}_p}(\{y \in \mathbb{Q}_p : f(y) \geq \lambda\}) \lesssim_d \lambda^{-\frac{d}{d-1}} |a_d|^{-\frac{1}{d-1}}.$$

A model example to keep in mind here is when $P(x) = a_d x^d$ is a monomial, in which case Ω consists of a single ball of radius $(r/|a_d|)^{1/d}$, with $P' = O_d(|a_d|(r/|a_d|)^{\frac{d-1}{d}})$ on this ball; also, one can verify $f(y) = O_d(|a_d|^{-1/d} r^{\frac{1}{d}-1})$ when $|y| \leq r$ and $f(y) = O_d(|a_d|^{-1/d}|y|^{\frac{1}{d}-1})$ when $|y| > r$. (The reader may wish to first verify these claims with $\mathbb{Q}_p$ replaced by $\mathbb{R}$ in order to build geometric intuition.) Note that this example also shows why all the exponents in the proposition are natural from a dimensional analysis (or scaling) perspective. Taking limits in (ii) as $r \to 0$, we also conclude that

$$\left\| \frac{dP_*\mu_{\mathbb{Q}_p}}{d\mu_{\mathbb{Q}_p}} \right\|_{L^{\frac{d}{d-1},\infty}(\mathbb{Q}_p)} \lesssim_d |a_d|^{-\frac{1}{d}},$$

where $\frac{dP_*\mu_{\mathbb{Q}_p}}{d\mu_{\mathbb{Q}_p}}$ is the Radon–Nikodym derivative (relative to Haar measure $\mu_{\mathbb{Q}_p}$) of the pushforward measure $P_*\mu_{\mathbb{Q}_p}$ of $\mu_{\mathbb{Q}_p}$ by P, and $L^{\frac{d}{d-1},\infty}$ is the weak $L^{\frac{d}{d-1}}$ norm; in the monomial case $P(x) = a_d x^d$, one can compute that this Radon–Nikodym derivative is proportional to the function $y \mapsto |a_d|^{-\frac{1}{d}}|y|^{\frac{1}{d}-1}$.

The van der Corput estimate in Proposition C.1(ii) can be also deduced from [52, Prop. 3.3. pp. 847], but for the convenience of the reader, we provide a self-contained proof.

Proof. To prove (i), we first work in the special case that P completely factorizes:

$$P(x) = c(x-\alpha_1)\cdots(x-\alpha_d)$$

for some $c, \alpha_1, \ldots, \alpha_d \in \mathbb{Q}_p$ with $c \neq 0$. We can cover Ω by $\Omega_1 \cup \ldots \cup \Omega_d$, where

$$\Omega_i \coloneqq \{x \in \Omega : |x-\alpha_i| \leq |x-\alpha_j| \text{ for all } j \in [d]\}.$$

It suffices to establish the claim (i) for a single Ω_i. Note from the ultratriangle inequality that for $x \in \Omega_i$ and $j \in [d]$, one has

$$|x-\alpha_j| = \max\{|x-\alpha_i|, |\alpha_i-\alpha_j|\},$$

and hence

$$|P(x)| = |c| \prod_{j=1}^{d} \max\{|x-\alpha_i|, |\alpha_i-\alpha_j|\}.$$

Thus we see that $\Omega_i \subseteq B(\alpha_i, R)$, where $R \in p^{\mathbb{Z}}$ is the maximal quantity for which

$$|c| \prod_{j=1}^{d} \max\{R, |\alpha_i - \alpha_j|\} \leq r.$$

On the other hand, we have from the product rule and triangle inequality for $x \in B(\alpha_i, R)$ that

$$\begin{aligned} |P'(x)| &\lesssim_d |c| \sup_{j\in[d]} \prod_{k\neq j} |x - \alpha_k| \\ &\lesssim_d |c| \sup_{j\in[d]} \prod_{k\neq j} \max\{R, |\alpha_i - \alpha_k|\} \\ &\lesssim_d R^{-1}|c| \prod_{k=1}^{d} \max\{R, |\alpha_i - \alpha_k|\} \\ &\lesssim_d \frac{r}{R}, \end{aligned}$$

giving Claim (i).

Now suppose that P only partially factorizes, thus

$$P(x) = (x - \alpha_1) \cdots (x - \alpha_j) Q(x)$$

for some $0 \leq j \leq d$ and some polynomial Q of degree $d - j$. The case $j = d$ has already been handled; now suppose inductively that $j < d$ and the claim (i) has already been proven for $j + 1$. We may assume Ω is non-empty since the claim (i) is trivial otherwise. Let α_{j+1} be an element of Ω that maximizes the magnitude of the quantity $\delta := (\alpha_{j+1} - \alpha_1) \cdots (\alpha_{j+1} - \alpha_j)$; such a quantity exists since Ω is compact, and δ is non-zero by continuity. Then

$$r \geq |P(\alpha_{j+1})| = |\delta||Q(\alpha_{j+1})|,$$

so $|Q(\alpha_{j+1})| \leq r/|\delta|$. By the factor theorem, we have

$$Q(x) = Q(\alpha_{j+1}) + (x - \alpha_{j+1})R(x)$$

for some polynomial R of degree $d - j - 1$, thus

$$P(x) = (x - \alpha_1) \cdots (x - \alpha_j) Q(\alpha_{j+1}) + (x - \alpha_1) \cdots (x - \alpha_{j+1}) R(x).$$

By construction, for $x \in \Omega$ we have $|P(x)| \leq r$, and

$$|(x - \alpha_1) \cdots (x - \alpha_j) Q(\alpha_{j+1})| \leq |\delta||Q(\alpha_{j+1})| \leq r,$$

hence by the ultratriangle inequality, we also have

$$|(x - \alpha_1) \cdots (x - \alpha_{j+1}) R(x)| \leq r.$$

By the induction hypothesis, we can cover Ω by $O_d(1)$ balls B on which the derivative of $(x - \alpha_1) \cdots (x - \alpha_{j+1}) R(x)$ is $O_d(r/\mu_{\mathbb{Q}_p}(B))$; by the $j = d$ case, we can also say the same about $(x - \alpha_1) \cdots (x - \alpha_j) Q(\alpha_{j+1})$. Intersecting the

balls together, we can say the same about P. This closes the induction and establishes the claim for any $0 \le j \le d$. Setting $j = 0$, we obtain (i).

Now we establish (ii). By iterating (i) d times and intersecting the balls together, we can cover Ω by $O_d(1)$ balls B on which $P^{(d)}(x) \lesssim_d r/\mu_{\mathbb{Q}_p}(B)^d$. But since $P^{(d)}(x) = d!a_d$, we have $\mu_{\mathbb{Q}_p}(B) \lesssim_d (r/|a_d|)^{1/d}$, giving the claim.

Now we prove (iii). Let $\lambda > 0$, and define the set

$$E \coloneqq \{y \in \mathbb{Q}_p : f(y) \ge \lambda\}.$$

Our task is to show that

$$\mu_{\mathbb{Q}_p}(E) \lesssim_d \lambda^{-\frac{d}{d-1}} |a_d|^{-\frac{1}{d-1}}.$$

If $y \in E$, then by definition

$$\mu_{\mathbb{Q}_p}(\{x \in \mathbb{Q}_p : |P(x) - y| \le r\}) \ge \lambda r.$$

By (i), the set in the left-hand side can be covered by $O_d(1)$ balls B, on which $|P'| \lesssim_d r/\mu_{\mathbb{Q}_p}(B)$. By the pigeonhole principle, one of these balls B must intersect the set in a set of measure $\gtrsim_d \lambda r$, thus $|P'| \lesssim_d r/(\lambda r) = 1/\lambda$ on this ball, and thus

$$\mu_{\mathbb{Q}_p}(\{x \in \mathbb{Q}_p : |P(x) - y| \le r \text{ and } |P'(x)| \lesssim_d 1/\lambda\}) \gtrsim_d \lambda r.$$

By the Fubini–Tonelli theorem, we conclude that

$$\mu_{\mathbb{Q}_p} \times \mu_{\mathbb{Q}_p}(\{(x,y) \in \mathbb{Q}_p^2 : |P(x) - y| \le r \text{ and } |P'(x)| \lesssim_d 1/\lambda\}) \gtrsim_d \lambda r \mu_{\mathbb{Q}_p}(E).$$

But by the Fubini–Tonelli theorem again, the left-hand side is equal to

$$r\mu_{\mathbb{Q}_p}(\{x \in \mathbb{Q}_p : |P'(x)| \lesssim_d 1/\lambda\}),$$

and hence by (ii) we obtain

$$\lambda r \mu_{\mathbb{Q}_p}(E) \lesssim_d r\Big(\frac{1}{\lambda |a_d|}\Big)^{\frac{1}{d-1}},$$

giving the claim. □

We can descend from the p-adics to a cyclic group of prime power order:

Corollary C.2 (Distribution of polynomials on a cyclic group of prime power order). *Let $Q = p^j$ for some $j \in \mathbb{Z}_+$, and let $P \in \mathbb{Z}[\mathrm{n}]$ be a polynomial of degree $d \ge 2$, which we also view as a map from $\mathbb{Z}/Q\mathbb{Z}$ to itself. Let $h\colon \mathbb{Z}/Q\mathbb{Z} \to \mathbb{N}$ be the counting function*

$$h(y) \coloneqq \#\{x \in \mathbb{Z}/Q\mathbb{Z} : P(x) = y\}.$$

Then for any $\lambda > 0$, we have the weak-type bound

$$\#\{y \in \mathbb{Z}/Q\mathbb{Z} : h(y) \ge \lambda\} \lesssim_P \lambda^{-\frac{d}{d-1}} Q.$$

In particular, one has

$$\|h\|_{L^s(\mathbb{Z}/Q\mathbb{Z})} \lesssim_{s,P} 1 \tag{C.3}$$

for any $0 < s < \frac{d}{d-1}$.

As before, the example of a monomial $P(x) = x^d$ shows that the range of s here is best possible. Interestingly, it seems difficult to establish this corollary without some version of the p-adic formalism, even though the statement of the corollary does not explicitly mention p-adics. Estimate (C.3) was previously obtained for monomials $P(x) = x^d$ in an unpublished work of Jim Wright on L^p-improving estimates for averaging operators on cyclic groups of the form $\mathbb{Z}/p^j\mathbb{Z}$ (private communication).

Proof. We can write $P(x) = a_d x^d + \cdots + a_0$, where $a_0, \ldots, a_d \in \mathbb{Z}_p$ are p-adic integers, thus they have norm at most 1. Note that

$$h(y) = Q\mu_{\mathbb{Q}_p}(\{x \in \mathbb{Q}_p : |x| \leq 1 \text{ and } |P(x) - y'| \leq Q^{-1}\})$$

for any $y \in \mathbb{Z}/Q\mathbb{Z}$ and $y' \in B(y, 1/Q)$, thus

$$\begin{aligned}\#\{y \in \mathbb{Z}/Q\mathbb{Z} : h(y) \geq \lambda\} \\ \leq Q\mu_{\mathbb{Q}_p}(\{y' \in \mathbb{Q}_p : Q\mu_{\mathbb{Q}_p}(\{x \in \mathbb{Q}_p : |P(x) - y'| \leq Q^{-1}\}) \geq \lambda\}).\end{aligned}$$

The claim now follows from Proposition C.1(iii). □

References

[1] T. Austin, A proof of Walsh's convergence theorem using couplings, *Int. Math. Res. Not. IMRN* no. 15 (2015), 6661–6674. MR 3384494. Zbl 1372.37012. https://doi.org/10.1093/imrn/rnu145.

[2] V. Bergelson, Weakly mixing PET, *Ergodic Theory Dynam. Systems* **7** no. 3 (1987), 337–349. MR 0912373. Zbl 0645.28012. https://doi.org/10.1017/S0143385700004090.

[3] V. Bergelson, Ergodic Ramsey theory—an update, in *Ergodic Theory of* $\mathbf{Z}^d$ *Actions* (Warwick, 1993–1994), *London Math. Soc. Lecture Note Ser.* **228**, Cambridge Univ. Press, Cambridge, 1996, pp. 1–61. MR 1411215. Zbl 0846.05095. https://doi.org/10.1017/CBO9780511662812.002.

[4] V. Bergelson, Combinatorial and Diophantine applications of ergodic theory, in *Handbook of Dynamical Systems. Vol.* 1*B*, Elsevier B. V., Amsterdam, 2006, (Appendix A by A. Leibman and Appendix B by Anthony Quas and Máté Wierdl), pp. 745–869. MR 2186252. Zbl 1130.37317. https://doi.org/10.1016/S1874-575X(06)80037-8.

[5] V. Bergelson and A. Leibman, Polynomial extensions of van der Waerden's and Szemerédi's theorems, *J. Amer. Math. Soc.* **9** no. 3 (1996), 725–753. MR 1325795. Zbl 0870.11015. https://doi.org/10.1090/S0894-0347-96-00194-4.

[6] V. Bergelson and A. Leibman, A nilpotent Roth theorem, *Invent. Math.* **147** no. 2 (2002), 429–470. MR 1881925. Zbl 1042.37001. https://doi.org/10.1007/s002220100179.

[7] G. D. Birkhoff, Proof of the ergodic theorem, *Proc. Natl. Acad. Sci. USA* **17** no. 12 (1931), 656–660. JFM 57.1011.02. https://doi.org/10.1073/pnas.17.2.656.

[8] M. Boshernitzan, G. Kolesnik, A. Quas, and M. Wierdl, Ergodic averaging sequences, *J. Anal. Math.* **95** (2005), 63–103. MR 2145587. Zbl 1100.28010. https://doi.org/10.1007/BF02791497.

[9] M. Boshernitzan and M. Wierdl, Ergodic theorems along sequences and Hardy fields, *Proc. Nat. Acad. Sci. U.S.A.* **93** no. 16 (1996), 8205–8207. MR 1401511. Zbl 0863.28011. https://doi.org/10.1073/pnas.93.16.8205.

[10] J. Bourgain, A nonlinear version of Roth's theorem for sets of positive density in the real line, *J. Analyse Math.* **50** (1988), 169–181. MR 0942826. Zbl 0675.42010. https://doi.org/10.1007/BF02796120.

[11] J. Bourgain, On the maximal ergodic theorem for certain subsets of the integers, *Israel J. Math.* **61** no. 1 (1988), 39–72. MR 0937581. Zbl 0642.28010. https://doi.org/10.1007/BF02776301.

[12] J. Bourgain, On the pointwise ergodic theorem on L^p for arithmetic sets, *Israel J. Math.* **61** no. 1 (1988), 73–84. MR 0937582. Zbl 0642.28011. https://doi.org/10.1007/BF02776302.

[13] J. Bourgain, Pointwise ergodic theorems for arithmetic sets, *Inst. Hautes Études Sci. Publ. Math.* no. 69 (1989), 5–45, with an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein. MR 1019960. Zbl 0705.28008. Available at http://www.numdam.org/item?id=PMIHES_1989__69__5_0.

[14] J. Bourgain, Double recurrence and almost sure convergence, *J. Reine Angew. Math.* **404** (1990), 140–161. MR 1037434. Zbl 0685.28008. https://doi.org/10.1515/crll.1990.404.140.

[15] J. Bourgain and M.-C. Chang, Nonlinear Roth type theorems in finite fields, *Israel J. Math.* **221** no. 2 (2017), 853–867. MR 3704938. Zbl 1420.11024. https://doi.org/10.1007/s11856-017-1577-9.

[16] F. Bruhat, Distributions sur un groupe localement compact et applications à l'étude des représentations des groupes $\wp$-adiques, *Bull. Soc. Math. France* **89** (1961), 43–75. MR 0140941. Zbl 0128.35701. https://doi.org/10.24033/bsmf.1559.

[17] Z. Buczolich and R. D. Mauldin, Divergent square averages, *Ann. of Math.* (2) **171** no. 3 (2010), 1479–1530. MR 2680392. Zbl 1200.37003. https://doi.org/10.4007/annals.2010.171.1479.

[18] A.-P. Calderón, Ergodic theory and translation-invariant operators, *Proc. Nat. Acad. Sci. U.S.A.* **59** (1968), 349–353. MR 0227354. Zbl 0185.21806. https://doi.org/10.1073/pnas.59.2.349.

[19] Q. Chu, N. Frantzikinakis, and B. Host, Ergodic averages of commuting transformations with distinct degree polynomial iterates, *Proc. Lond. Math. Soc.*

(3) **102** no. 5 (2011), 801–842. MR 2795725. Zbl 1218.37009. https://doi.org/10.1112/plms/pdq037.

[20] S. Dasu, C. Demeter, and B. Langowski, Sharp ℓ^p-improving estimates for the discrete paraboloid, *J. Fourier Anal. Appl.* **27** no. 1 (2021), Paper No. 3. MR 4199855. Zbl 1457.42013. https://doi.org/10.1007/s00041-020-09801-2.

[21] C. Demeter, Pointwise convergence of the ergodic bilinear Hilbert transform, *Illinois J. Math.* **51** no. 4 (2007), 1123–1158. MR 2417419. Zbl 1175.37013. https://doi.org/10.1215/ijm/1258138536.

[22] C. Demeter, M. T. Lacey, T. Tao, and C. Thiele, Breaking the duality in the return times theorem, *Duke Math. J.* **143** no. 2 (2008), 281–355. MR 2420509. Zbl 1213.42064. https://doi.org/10.1215/00127094-2008-020.

[23] C. Demeter, T. Tao, and C. Thiele, Maximal multilinear operators, *Trans. Amer. Math. Soc.* **360** no. 9 (2008), 4989–5042. MR 2403711. Zbl 1268.42034. https://doi.org/10.1090/S0002-9947-08-04474-7.

[24] J.-M. Derrien and E. Lesigne, Un théorème ergodique polynomial ponctuel pour les endomorphismes exacts et les K-systèmes, *Ann. Inst. H. Poincaré Probab. Statist.* **32** no. 6 (1996), 765–778. MR 1422310. Zbl 0868.60028. Available at http://www.numdam.org/item/AIHPB_1996__32_6_765_0/.

[25] Y. Do, C. Muscalu, and C. Thiele, Variational estimates for paraproducts, *Rev. Mat. Iberoam.* **28** no. 3 (2012), 857–878. MR 2949622. Zbl 1256.42018. https://doi.org/10.4171/RMI/694.

[26] D. Dong, On a discrete bilinear singular operator, *C. R. Math. Acad. Sci. Paris* **355** no. 5 (2017), 538–542. MR 3650379. Zbl 1365.42009. https://doi.org/10.1016/j.crma.2017.03.010.

[27] D. Dong, X. Li, and W. Sawin, Improved estimates for polynomial Roth type theorems in finite fields, *J. Anal. Math.* **141** no. 2 (2020), 689–705. MR 4179774. Zbl 1475.11022. https://doi.org/10.1007/s11854-020-0113-8.

[28] P. Durcik, S. Guo, and J. Roos, A polynomial Roth theorem on the real line, *Trans. Amer. Math. Soc.* **371** no. 10 (2019), 6973–6993. MR 3939567. Zbl 1432.42005. https://doi.org/10.1090/tran/7574.

[29] el H. el Abdalaoui, Simple proof of Bourgain bilinear ergodic theorem and its extension to polynomials and polynomials in primes, 2019. arXiv 1908.02281.

[30] N. Frantzikinakis, Some open problems on multiple ergodic averages, *Bull. Hellenic Math. Soc.* **60** (2016), 41–90. MR 3613710. Zbl 1425.37004. arXiv 11903,3808.

[31] N. Frantzikinakis and B. Kra, Polynomial averages converge to the product of integrals. Probability in mathematics, *Israel J. Math.* **148** (2005), 267–276. MR 2191231. Zbl 1155.37303. https://doi.org/10.1007/BF02775439.

[32] N. Frantzikinakis, E. Lesigne, and M. Wierdl, Random differences in Szemerédi's theorem and related results, *J. Anal. Math.* **130** (2016), 91–133. MR 3574649. Zbl 1414.11018. https://doi.org/10.1007/s11854-016-0030-z.

[33] H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. Analyse Math.* **31** (1977), 204–256. MR 0498471. Zbl 0347.28016. https://doi.org/10.1007/BF02813304.

[34] H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton Univ. Press, Princeton, N.J., 1981, M. B. Porter Lectures. MR 0603625. Zbl 0459.28023. https://doi.org/10.1515/9781400855162.

[35] H. Furstenberg, Nonconventional ergodic averages, in *The Legacy of John von Neumann* (Hempstead, NY, 1988), *Proc. Sympos. Pure Math.* **50**, Amer. Math. Soc., Providence, RI, 1990, pp. 43–56. MR 1067751. Zbl 0711.28006. https://doi.org/10.1090/pspum/050/1067751.

[36] H. Furstenberg and B. Weiss, A mean ergodic theorem for $(1/N)\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$, in *Convergence in Ergodic Theory and Probability* (Columbus, OH, 1993), *Ohio State Univ. Math. Res. Inst. Publ.* **5**, de Gruyter, Berlin, 1996, pp. 193–227. MR 1412607. Zbl 0869.28010. https://doi.org/10.1515/9783110889383.193.

[37] A. Gaitan and V. Lie, The boundedness of the (sub)bilinear maximal function along "non-flat" smooth curves, *J. Fourier Anal. Appl.* **26** no. 4 (2020), Paper No. 69, 33. MR 4131966. Zbl 1447.42018. https://doi.org/10.1007/s00041-020-09770-6.

[38] W. T. Gowers, Decompositions, approximate structure, transference, and the Hahn-Banach theorem, *Bull. Lond. Math. Soc.* **42** no. 4 (2010), 573–606. MR 2669681. Zbl 1233.05198. https://doi.org/10.1112/blms/bdq018.

[39] W. T. Gowers and J. Wolf, The true complexity of a system of linear equations, *Proc. Lond. Math. Soc.* (3) **100** no. 1 (2010), 155–176. MR 2578471. Zbl 1243.11010. https://doi.org/10.1112/plms/pdp019.

[40] B. Green and T. Tao, Quadratic uniformity of the Möbius function, *Ann. Inst. Fourier* (*Grenoble*) **58** no. 6 (2008), 1863–1935. MR 2473624. Zbl 1160.11017. https://doi.org/10.5802/aif.2401.

[41] R. Han, V. Kovač, M. T. Lacey, J. Madrid, and F. Yang, Improving estimates for discrete polynomial averages, *J. Fourier Anal. Appl.* **26** no. 3 (2020), Paper No. 42, 11. MR 4106792. Zbl 1442.42024. https://doi.org/10.1007/s00041-020-09748-4.

[42] E. Hopf, The general temporally discrete Markoff process, *J. Rational Mech. Anal.* **3** (1954), 13–45. MR 0060181. Zbl 0055.36705. https://doi.org/10.1512/iumj.1954.3.53002.

[43] B. Host and B. Kra, Convergence of polynomial ergodic averages. Probability in mathematics, *Israel J. Math.* **149** (2005), 1–19. MR 2191208. Zbl 1085.28009. https://doi.org/10.1007/BF02772534.

[44] B. Host and B. Kra, Nonconventional ergodic averages and nilmanifolds, *Ann. of Math.* (2) **161** no. 1 (2005), 397–488. MR 2150389. Zbl 1077.37002. https://doi.org/10.4007/annals.2005.161.397.

[45] B. Host and B. Kra, A point of view on Gowers uniformity norms, *New York J. Math.* **18** (2012), 213–248. MR 2920990. Zbl 1268.11018. Available at http://nyjm.albany.edu:8000/j/2012/18_213.html.

[46] A. D. Ionescu, Á. Magyar, M. Mirek, and T. Z. Szarek, Polynomial averages and pointwise ergodic theorems on nilpotent groups, 2021. arXiv 2112.03322.

[47] A. D. Ionescu and S. Wainger, L^p boundedness of discrete singular Radon transforms, *J. Amer. Math. Soc.* **19** no. 2 (2006), 357–383. MR 2188130. Zbl 1158.42007. https://doi.org/10.1090/S0894-0347-05-00508-4.

[48] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, *Amer. Math. Soc. Colloq. Publ.* **53**, Amer. Math. Soc., Providence, RI, 2004. MR 2061214. Zbl 1059.11001. https://doi.org/10.1090/coll/053.

[49] R. L. Jones, R. Kaufman, J. M. Rosenblatt, and M. Wierdl, Oscillation in ergodic theory, *Ergodic Theory Dynam. Systems* **18** no. 4 (1998), 889–935. MR 1645330. Zbl 0924.28009. https://doi.org/10.1017/S0143385798108349.

[50] R. L. Jones, A. Seeger, and J. Wright, Strong variational and jump inequalities in harmonic analysis, *Trans. Amer. Math. Soc.* **360** no. 12 (2008), 6711–6742. MR 2434308. Zbl 1159.42013. https://doi.org/10.1090/S0002-9947-08-04538-8.

[51] R. L. Jones and G. Wang, Variation inequalities for the Fejér and Poisson kernels, *Trans. Amer. Math. Soc.* **356** no. 11 (2004), 4493–4518. MR 2067131. Zbl 1065.42006. https://doi.org/10.1090/S0002-9947-04-03397-5.

[52] M. W. Kowalski and J. Wright, Elementary inequalities involving the roots of a polynomial with applications in harmonic analysis and number theory, *J. Lond. Math. Soc.* (2) **86** no. 3 (2012), 835–851. MR 3000832. Zbl 1291.11109. https://doi.org/10.1112/jlms/jds029.

[53] B. Kra, Private communication, October (2021).

[54] B. Krause, Polynomial ergodic averages converge rapidly: Variations on a theorem of Bourgain, 2014, to appear in *Israel J. Math.* arXiv 1402.1803.

[55] M. T. Lacey, The bilinear maximal functions map into L^p for $2/3 < p \leq 1$, *Ann. of Math.* (2) **151** no. 1 (2000), 35–57. MR 1745019. Zbl 0967.47031. https://doi.org/10.2307/121111.

[56] P. LaVictoire, Universally L^1-bad arithmetic sequences, *J. Anal. Math.* **113** (2011), 241–263. MR 2788358. Zbl 1220.37005. https://doi.org/10.1007/s11854-011-0006-y.

[57] A. Leibman, Convergence of multiple ergodic averages along polynomials of several variables, *Israel J. Math.* **146** (2005), 303–315. MR 2151605. Zbl 1080.37002. https://doi.org/10.1007/BF02773538.

[58] A. Leibman, Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold, *Ergodic Theory Dynam. Systems* **25** no. 1 (2005), 201–213. MR 2122919. Zbl 1080.37003. https://doi.org/10.1017/S0143385704000215.

[59] A. Lewko and M. Lewko, Estimates for the square variation of partial sums of Fourier series and their rearrangements, *J. Funct. Anal.* **262** no. 6 (2012), 2561–2607. MR 2885959. Zbl 1239.42027. https://doi.org/10.1016/j.jfa.2011.12.007.

[60] X. Li, Bilinear Hilbert transforms along curves I: The monomial case, *Anal. PDE* **6** no. 1 (2013), 197–220. MR 3068544. Zbl 1275.42022. https://doi.org/10.2140/apde.2013.6.197.

[61] X. Li and L. Xiao, Uniform estimates for bilinear Hilbert transforms and bilinear maximal functions associated to polynomials, *Amer. J. Math.* **138** no. 4 (2016), 907–962. MR 3538147. Zbl 1355.42011. https://doi.org/10.1353/ajm.2016.0030.

[62] V. Lie, On the boundedness of the bilinear Hilbert transform along "non-flat" smooth curves, *Amer. J. Math.* **137** no. 2 (2015), 313–363. MR 3337797. Zbl 1318.42008. https://doi.org/10.1353/ajm.2015.0013.

[63] V. Lie, On the boundedness of the bilinear Hilbert transform along "non-flat" smooth curves. The Banach triangle case $(L^r, 1 \leq r < \infty)$, *Rev. Mat. Iberoam.* **34** no. 1 (2018), 331–353. MR 3763348. Zbl 1395.42009. https://doi.org/10.4171/RMI/987.

[64] Á. Magyar, E. M. Stein, and S. Wainger, Discrete analogues in harmonic analysis: spherical averages, *Ann. of Math.* (2) **155** no. 1 (2002), 189–208. MR 1888798. Zbl 1036.42018. https://doi.org/10.2307/3062154.

[65] M. Mirek, Square function estimates for discrete Radon transforms, *Anal. PDE* **11** no. 3 (2018), 583–608. MR 3738256. Zbl 1383.42014. https://doi.org/10.2140/apde.2018.11.583.

[66] M. Mirek, E. M. Stein, and B. Trojan, $\ell^p(\mathbb{Z}^d)$-estimates for discrete operators of Radon type: variational estimates, *Invent. Math.* **209** no. 3 (2017), 665–748. MR 3681393. Zbl 1397.42006. https://doi.org/10.1007/s00222-017-0718-4.

[67] M. Mirek, E. M. Stein, and P. Zorin-Kranich, A bootstrapping approach to jump inequalities and their applications, *Anal. PDE* **13** no. 2 (2020), 527–558. MR 4078235. Zbl 1437.42023. https://doi.org/10.2140/apde.2020.13.527.

[68] M. Mirek, E. M. Stein, and P. Zorin-Kranich, Jump inequalities for translation-invariant operators of Radon type on $\mathbb{Z}^d$, *Adv. Math.* **365** (2020), 107065, 57. MR 4067360. Zbl 1436.42022. https://doi.org/10.1016/j.aim.2020.107065.

[69] M. Mirek, E. M. Stein, and P. Zorin-Kranich, Jump inequalities via real interpolation, *Math. Ann.* **376** no. 1-2 (2020), 797–819. MR 4055178. Zbl 1448.60052. https://doi.org/10.1007/s00208-019-01889-2.

[70] J. von Neumann, Proof of the quasi-ergodic hypothesis, *Proc. Natl. Acad. Sci. USA* **18** (1932), 70–82. Zbl 0004.31004. https://doi.org/10.1073/pnas.18.1.70.

[71] M. S. Osborne, On the Schwartz-Bruhat space and the Paley-Wiener theorem for locally compact abelian groups, *J. Functional Analysis* **19** (1975), 40–49. MR 0407534. Zbl 0295.43008. https://doi.org/10.1016/0022-1236(75)90005-1.

[72] S. Peluse, Three-term polynomial progressions in subsets of finite fields, *Israel J. Math.* **228** no. 1 (2018), 379–405. MR 3874848. Zbl 1456.11015. https://doi.org/10.1007/s11856-018-1768-z.

[73] S. Peluse, *Bounds for Sets with No Nontrivial Polynomial Progressions*, ProQuest LLC, Ann Arbor, MI, 2019, Thesis (Ph.D.)–Stanford Univ. MR 4197632.

[74] S. Peluse and S. Prendiville, Quantitative bounds in the non-linear Roth theorem, 2019. arXiv 1903.02592.

[75] S. Peluse and S. Prendiville, A polylogarithmic bound in the nonlinear Roth theorem, *Int. Math. Res. Not. IMRN* no. 10 (2020). https://doi.org/10.1093/imrn/rnaa261.

[76] L. B. Pierce, On superorthogonality, *J. Geom. Anal.* **31** no. 7 (2021), 7096–7183, with an appendix by Emmanuel Kowalski. MR 4289256. Zbl 1471.42037. https://doi.org/10.1007/s12220-021-00606-3.

[77] S. Prendiville, The inverse theorem for the nonlinear Roth configuration: an exposition, 2020. arXiv 2003.04121.

[78] W. Rudin, *Fourier Analysis on Groups*, *Interscience Tracts Pure Appl. Math.* **12**, Interscience Publ. (a division of John Wiley and Sons), New York, 1962. MR 0152834. Zbl 1361.43001. https://doi.org/10.1002/9781118165621.

[79] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 199–245. MR 0369312. Zbl 0335.10054. https://doi.org/10.4064/aa-27-1-199-245.

[80] T. Tao, The Ionescu-Wainger multiplier theorem and the adeles, *Mathematika* **67** no. 3 (2021), 647–677. MR 4270396. https://doi.org/10.1112/mtk.12094.

[81] J. T. Tate, Fourier analysis in number fields, and Hecke's zeta-functions, in *Algebraic Number Theory* (Proc. Instructional Conf., Brighton, 1965), Thompson, Washington, D.C., 1967, pp. 305–347. MR 0217026.

[82] M. N. Walsh, Norm convergence of nilpotent ergodic averages, *Ann. of Math.* (2) **175** no. 3 (2012), 1667–1688. MR 2912715. Zbl 1248.37008. https://doi.org/10.4007/annals.2012.175.3.15.

[83] T. Ziegler, Universal characteristic factors and Furstenberg averages, *J. Amer. Math. Soc.* **20** no. 1 (2007), 53–97. MR 2257397. Zbl 1198.37014. https://doi.org/10.1090/S0894-0347-06-00532-7.

(Received: August 1, 2020)
(Revised: December 20, 2021)

Princeton University, Princeton, NJ, USA
Current address: King's College London, London, UK
E-mail: ben.krause@kcl.ac.uk

Rutgers University, Piscataway, NJ, USA and
Uniwersytet Wrocławski, Wrocław, Poland
E-mail: mariusz.mirek@rutgers.edu

University of California Los Angeles, Los Angeles, CA, USA
E-mail: tao@math.ucla.edu