

Pointwise convergence of bilinear polynomial averages over the primes

BEN KRAUSE[†], HAMED MOUSAVI[†], TERENCE TAO[‡] and
JONI TERÄVÄINEN[§]

[†] *Department of Mathematics, University of Bristol, Bristol BS8 1QU, UK*
(e-mail: ben.krause@bristol.ac.uk, gj23799@bristol.ac.uk)

[‡] *Department of Mathematics, University of California, Los Angeles,
CA 90095-1555, USA*
(e-mail: tao@math.ucla.edu)

[§] *Department of Pure Mathematics and Mathematical Statistics,
University of Cambridge, Cambridge CB3 0WB, UK*
(e-mail: joni.p.teravainen@gmail.com)

(Received 2 October 2024 and accepted in revised form 18 June 2025)

Abstract. We show that on a σ -finite measure-preserving system $X = (X, \nu, T)$, the non-conventional ergodic averages

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

converge pointwise almost everywhere for $f \in L^{p_1}(X)$, $g \in L^{p_2}(X)$ and $1/p_1 + 1/p_2 \leq 1$, where P is a polynomial with integer coefficients of degree at least 2. This had previously been established with the von Mangoldt weight Λ replaced by the constant weight 1 by the first and third authors with Mirek, and by the Möbius weight μ by the fourth author. The proof is based on combining tools from both of these papers, together with several Gowers norm and polynomial averaging operator estimates on approximants to the von Mangoldt function of ‘Cramér’ and ‘Heath-Brown’ type.

Key words: pointwise convergence, bilinear ergodic averages, Gowers norms
2020 Mathematics Subject Classification: 37A30 (Primary); 37A44, 37A46, 11B30 (Secondary)

1. Introduction

Throughout this paper, $P \in \mathbb{Z}[n]$ denotes a polynomial with integer coefficients of some degree $d \geq 2$ in one indeterminate n ; a typical case to keep in mind is the quadratic polynomial $P(n) = n^2$.

Define a *measure-preserving system* to be a triple $X = (X, \nu, T)$, where $X = (X, \nu)$ is a σ -finite measure space and $T: X \rightarrow X$ is an invertible bimeasurable map which is measure-preserving in the sense that $\nu(T^{-1}(E)) = \nu(E)$ for all measurable E . It is common in the literature to restrict to finite measure systems and to normalize $\nu(X) = 1$; but our results will not require any hypothesis of finite measure. Given functions $f, g: X \rightarrow \mathbb{C}$, a scale $N \geq 1$ and a weight function $w: \mathbb{N} \rightarrow \mathbb{C}$, we can then define the non-conventional averaging operator

$$A_{N,w;X}(f, g)(x) := \mathbb{E}_{n \in [N]} w(n) f(T^n x) g(T^{P(n)} x)$$

for any $x \in X$ (see §2 for our averaging notation).

1.1. Unweighted ergodic averages. In the unweighted case $w = 1$, the following ergodic theorem was recently proven by two of the authors with Mirek.

THEOREM 1.1. (Unweighted ergodic theorem [13, Theorem 1.17]) *Let (X, ν, T) be a measure-preserving system and let $f \in L^{p_1}(X)$, $g \in L^{p_2}(X)$ for some $1 < p_1, p_2 < \infty$ with $(1/p_1) + (1/p_2) = (1/p) \leq 1$.*

- (i) *(Mean ergodic theorem) The averages $A_{N,1;X}(f, g)$ converge in $L^p(X)$ norm.*
- (ii) *(Pointwise ergodic theorem) The averages $A_{N,1;X}(f, g)$ converge pointwise almost everywhere.*
- (iii) *(Maximal ergodic theorem) One has*

$$\|(A_{N,1;X}(f, g))_{N \in \mathbb{Z}^+}\|_{L^p(X; \ell^\infty)} \lesssim_{p_1, p_2, p} \|f\|_{L^{p_1}(X)} \|g\|_{L^{p_2}(X)}$$

(see §2.2 for our asymptotic notation conventions).

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

$$\|(A_{N,1;X}(f, g))_{N \in \mathbb{D}}\|_{L^p(X; \mathbf{V}^r)} \lesssim_{p_1, p_2, r, p, \lambda} \|f\|_{L^{p_1}(X)} \|g\|_{L^{p_2}(X)}$$

whenever $\mathbb{D} \subset [1, +\infty)$ is finite and λ -lacunary (see §2.6 for the definition of λ -lacunarity and the variational norm \mathbf{V}^r).

We very briefly review the main ingredients of the proof of Theorem 1.1. Case (iv) is the main estimate, which easily implies the other three claims. By some standard sparsification and transference arguments, as well as dyadic decompositions, it sufficed to prove the variant estimate

$$\|(\tilde{A}_{N,1}(f, g))_{N \in \mathbb{D}}\|_{\ell^p(\mathbb{Z}; \mathbf{V}^r)} \lesssim_{p_1, p_2, r, p, \lambda} \|f\|_{\ell^{p_1}(\mathbb{Z})} \|g\|_{\ell^{p_2}(\mathbb{Z})},$$

where

$$\tilde{A}_{N,w}(f, g)(x) := \mathbb{E}_{n \in [N]} w(n) f(x+n) g(x+P(n)) \mathbb{1}_{n > N/2} \quad (1.1)$$

is the ‘upper half’ of $A_{N,w;X}$ when X is the integers \mathbb{Z} with the usual shift $T: n \mapsto n+1$ and counting measure ν .

A crucial observation was that the averages $\tilde{A}_{N,1}$ are ‘complexity zero’ in the sense that they are small when the Fourier transform of f or g vanish on ‘major arcs’. Indeed, in [13, Theorem 5.12], the single-scale minor arc estimate

$$\|\tilde{A}_{N,1}(f, g)\|_{\ell^1(\mathbb{Z})} \lesssim_{C_1} (2^{-cl} + (\text{Log } N)^{-cC_1}) \|f\|_{\ell^2(\mathbb{Z})} \|g\|_{\ell^2(\mathbb{Z})} \quad (1.2)$$

was proven for $N \geq 1, l \in \mathbb{N}$ and $f, g \in \ell^2(\mathbb{Z})$ with either the Fourier transform $\mathcal{F}_{\mathbb{Z}} f$ of f vanishing on the major arc set $\mathcal{M}_{\leq l, \leq -\text{Log } N + l}$ or the Fourier transform $\mathcal{F}_{\mathbb{Z}} g$ of g vanishing on the major arc set $\mathcal{M}_{\leq l, \leq -d \text{Log } N + dl}$; we refer the reader to §2 for the definition of the various terms and symbols introduced here. This minor arc estimate was proven by combining Peluse–Prendiville estimates [24] with a discrete ℓ^p improving inequality from [8], together with a Hahn–Banach argument.

Using equation (1.2), one could now focus attention to major arcs. After some routine manipulations involving Ionescu–Wainger multiplier theory [10], the task reduced to controlling the $\ell^p(\mathbb{Z}; \mathbf{V}^r)$ norm of tuples of the form

$$(\tilde{A}_{N,1}(F_N, G_N))_{N \in \mathbb{I}}, \quad (1.3)$$

where \mathbb{I} is a certain λ -lacunary set (bounded from below by certain bounds, but not from above) and F_N, G_N are various frequency localizations of f, g , respectively, to major arcs (see [13, Theorem 5.30] for a precise statement). By estimation of the bilinear symbol of the averaging operator $\tilde{A}_{N,1}$, one could approximate this tuple by another tuple

$$(B_{(\varphi_N \otimes \tilde{\varphi}_N) \tilde{m}_{N, \mathbb{R}}}^{l_1, l_2, m_{\hat{\mathbb{Z}}}}(F, G))_{N \in \mathbb{I}}, \quad (1.4)$$

where F, G are again some Fourier localizations of f, g to major arcs and $B_{(\varphi_N \otimes \tilde{\varphi}_N) \tilde{m}_{N, \mathbb{R}}}^{l_1, l_2, m_{\hat{\mathbb{Z}}}}$ is a certain bilinear Fourier multiplier adapted to major arcs; see [13, Proposition 7.13] for a precise statement. At this stage, it became necessary to split the set \mathbb{I} of spatial averaging scales into the small scales \mathbb{I}_{\leq} and large scales $\mathbb{I}_{>}$. For the small scales, one could reduce matters to controlling another tuple

$$(B_{m_*}^{l_1, l_2, m_{\hat{\mathbb{Z}}}}(T_{\varphi_{N, \mathfrak{f}, j_1}}^{l_1} F, T_{\tilde{\varphi}_{N, \mathfrak{f}, j_2}}^{l_2} G))_{N \in \mathbb{I}_{\leq}}$$

for another bilinear Fourier multiplier $B_{m_*}^{l_1, l_2, m_{\hat{\mathbb{Z}}}}$ and Fourier multipliers $T_{\varphi_{N, \mathfrak{f}, j_1}}^{l_1}, T_{\tilde{\varphi}_{N, \mathfrak{f}, j_2}}^{l_2}$, while for the large scales, one instead considered tuples of the form

$$(B_{1 \otimes m_{\hat{\mathbb{Z}}}}(T_{\varphi_{N, \mathfrak{f}, j_1}} \otimes 1 F_{\mathbb{A}}, T_{\tilde{\varphi}_{N, \mathfrak{f}, j_2}} \otimes 1 G_{\mathbb{A}}))_{N \in \mathbb{I}_{>}},$$

where $F_{\mathbb{A}}, G_{\mathbb{A}}$ were now defined on the ring $\mathbb{A}_{\mathbb{Z}} = \mathbb{R} \times \hat{\mathbb{Z}}$ of adelic integers rather than on the integers \mathbb{Z} . See [13, Theorem 7.28] for a precise statement of the estimates required on these tuples.

In the small-scale case, it was possible to apply a general two-parameter Radamacher–Menshov inequality [13, Corollary 8.2] followed by some shifted Calderón–Zygmund theory [13, Theorem B.1] to reduce matters to obtaining a good $\ell^{p_1}(\mathbb{Z}) \times \ell^{p_2}(\mathbb{Z}) \rightarrow \ell^p(\mathbb{Z})$ estimate for the bilinear multiplier $B_{m_*}^{l_1, l_2, m_{\hat{\mathbb{Z}}}}$ (see [13, Lemma 8.6]), which was ultimately proven with the assistance of the minor arc estimate in equation (1.2) and the approximation result in [13, Proposition 7.13].

In the large-scale case, some interpolation and factorization arguments, together with a version of equation (1.2) on the profinite integers $\hat{\mathbb{Z}}$, reduced matters to establishing $L^2(\mathbb{Z}_p) \times L^2(\mathbb{Z}_p) \rightarrow L^q(\mathbb{Z}_p)$ bounds on the p -adic averaging operator

$$A_{\mathbb{Z}_p}(f, g)(x) := \mathbb{E}_{n \in \mathbb{Z}_p} f(x+n)g(x+P(n)) \quad (1.5)$$

for all primes p and some $q > 2$, with the operator norm required to be bounded by 1 for p large enough; see [13, equations (10.3), (10.4)] for a precise statement. The boundedness ultimately came from some distributional analysis of the level sets of P on the p -adics (see [13, Corollary C.2]); getting the bound of 1 for large p required some additional refined analysis in which one again uses (a p -adic version of) the minor arc estimate in equation (1.2).

1.2. Möbius-weighted ergodic averages. More recently, another one of us [26] considered the non-conventional averaging operators $A_{N,\mu;X}$ weighted by the Möbius function μ instead of 1. Perhaps counter-intuitively, the convergence of ergodic averages weighted by μ is actually *better* than that of the unweighted case, especially in light of the recent progress on quantitative Gowers uniformity of the Möbius function [7, 14–16, 25]. For instance, as a special case of [26, Theorem 1.2], the following result was shown.

THEOREM 1.2. (Möbius-weighted ergodic theorem) *Let X have finite measure, $f \in L^{p_1}(X)$, $g \in L^{p_2}(X)$ with $(1/p_1) + (1/p_2) < 1$, and let $A > 0$. Then,*

$$\lim_{N \rightarrow \infty} (\log^A N) A_{N,\mu;X}(f, g) = 0 \quad (1.6)$$

pointwise almost everywhere.

The ingredients used to prove Theorem 1.2 are somewhat different from those used to prove Theorem 1.1; a key input was [26, Theorem 4.1], which, in our context, establishes the bound

$$|\mathbb{E}_{x \in [-CN^d, CN^d]} A_{N,\theta;\mathbb{Z}}(f, g)(x)h(x)| \lesssim_{C,P} (N^{-1} + \|\theta\|_{u^{d+1}[N]})^{1/K} \quad (1.7)$$

for all 1-bounded f, g, h, θ and some $1 \leq K \lesssim_d 1$, where the ‘little’ Gowers uniformity norm $\|\theta\|_{u^{d+1}[N]}$ is defined as

$$\|\theta\|_{u^{d+1}[N]} := \sup_{\deg Q \leq d} |\mathbb{E}_{n \in [N]} \theta(n)e(-Q(n))|, \quad (1.8)$$

where Q ranges over all polynomials of degree at most d with real coefficients and $e(x) := e^{2\pi i x}$. The results of [7] show that $\|\mu\|_{u^{d+1}[N]}$ decays faster than any power of $\log N$, and the claim then follows by standard sparsification and transference arguments.

1.3. Prime-weighted ergodic averages. In this paper, we combine the methods of [13, 26], together with some additional arguments, to obtain a non-conventional ergodic theorem in which the weight is selected to be the von Mangoldt function Λ , defined by

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

More specifically, we show the following.

THEOREM 1.3. (Main theorem) *Let (X, ν, T) be a measure-preserving system and let $f \in L^{p_1}(X)$, $g \in L^{p_2}(X)$ for some $1 < p_1, p_2 < \infty$ with $(1/p_1) + (1/p_2) \leq 1$. Then, the averages $A_{N, \Lambda; X}(f, g)$ converge pointwise almost everywhere. In fact, one has the variational estimate*

$$\|(A_{N, \Lambda; X}(f, g))_{N \in \mathbb{D}}\|_{L^p(X; \mathbf{V}^r)} \lesssim_{p_1, p_2, p, r, P, \lambda} \|f\|_{L^{p_1}(X)} \|g\|_{L^{p_2}(X)} \quad (1.9)$$

whenever $\lambda > 1$, $p \geq 1$ and $r > 2$ with $(1/p_1) + (1/p_2) = (1/p)$, and $\mathbb{D} \subset [1, +\infty)$ is finite and λ -lacunary.

The range of r here is optimal, as will be mentioned in §6.4. It is possible to extend the range of (p_1, p_2) slightly beyond duality, see the discussion in §6.3.

Using the fact that $\log n = \log N + O(\log M)$ for $n \in [N/M, N]$ and the prime number theorem, we have the following immediate corollary to Theorem 1.3.

COROLLARY 1.4. *Let the assumptions be as in Theorem 1.3. Then, the prime-weighted averages*

$$\frac{1}{N/\log N} \sum_{p \leq N} f(T^p x) g(T^{P(p)} x)$$

converge pointwise almost everywhere.

Previously, the pointwise convergence of ergodic averages over the primes was known only in the case of a single polynomial iterate. This case was established by Bourgain [2] and Wierdl [27] for linear polynomials (with the latter work allowing L^q functions for any $q > 1$), and the case of an arbitrary single polynomial iterate was handled by Nair [21, 22]. We also mention that the problem of pointwise convergence of ergodic averages with more than one iterate was discussed by Frantzikinakis in [3, Problem 12]; the specific problem there about two linear iterates however remains open.

Let us also mention that the *norm convergence* of non-conventional ergodic averages is now known for any number of polynomial iterates, thanks to the works of Frantzikinakis, Host and Kra [4], and Wooley and Ziegler [28].

1.4. Methods of proof. From a high-level perspective, Theorem 1.3 is proven by combining the methods used in [13] to prove Theorem 1.1 with the methods used in [26] to prove Theorem 1.2. However, several technical difficulties make the analysis delicate in places, as we shall now discuss.

The first issue arises when trying to approximate various frequency-localized averages (analogous to equation (1.3), but with the weight 1 replaced by Λ) by certain bilinear model operators (analogous to equation (1.4), but with the symbol $m_{\hat{\mathbb{Z}}}$ replaced by a variant $m_{\hat{\mathbb{Z}} \times}$). It is important for the arguments in [13] that the error in this approximation gains a polynomial factor N^{-c} in N , or at least a quasipolynomial factor $\exp(-\log^c N)$. Using the von Mangoldt function as a weight, this is possible in the absence of Siegel zeroes (and, in particular, assuming the generalized Riemann hypothesis); however, the presence of a Siegel zero near a given scale N requires one to add a scale-dependent correction term to the bilinear symbol $m_{\hat{\mathbb{Z}}}$ to obtain a satisfactory approximation at small scales.

While this correction term is ultimately manageable because of the Landau–Page theorem, it significantly complicates the analysis, in that one cannot simply repeat arguments from [13] verbatim. See §6 for further discussion.

To avoid this issue, we adapt some ideas from [25] and swap the von Mangoldt weight Λ early in the argument with an approximant Λ_N that is not sensitive to Siegel zeroes. The arguments used in [26] to establish Theorem 1.2 allow one to do so provided that one has good control of the little Gowers uniformity norm in the sense that

$$\|\Lambda - \Lambda_N\|_{u^{d+1}[N]} \lesssim (\log N)^{-A}$$

for some large A . One available choice of approximant is the *Cramér(–Granville) approximant*

$$\Lambda_{\text{Cramér},w}(n) := \frac{W}{\varphi(W)} \mathbb{1}_{(n,W)=1}$$

for a suitable parameter w and $W = \prod_{p \leq w} p$ (we end up selecting $w := \exp(\log^{1/C_0} N)$ for some large constant C_0); the required bounds follow, for instance, from the results in [18] (which even extend to shorter intervals). A useful fact, first observed in [25] and refined further here, is that these approximants are stable in Gowers uniformity norms with respect to the w parameter; see Lemma 4.5 for a precise statement.

After using the arguments from [26] to replace Λ by Λ_N , most of the arguments of [13] proceed with only minor changes; in particular, the analogue of the approximation of equation (1.3) by equation (1.4) is fairly routine, thanks in large part to the fundamental lemma of sieve theory; see the proof of Proposition 3.4 in §5. We remark that Siegel zeroes play no role whatsoever in establishing this proposition, in contrast to what would have occurred if we retained the original weight Λ instead of Λ_N . However, three components of the argument of Theorem 1.3 still require some additional care. The first is a *polynomial improving estimate*

$$\left(\sum_{x \in \mathbb{Z}} |\mathbb{E}_{n \in [N]} (\Lambda(n) + \Lambda_N(n)) f(x + P(n))|^2 \right)^{1/2} \lesssim N^{d(1/2-1/p)} \|f\|_{\ell^p(\mathbb{Z})}$$

for $p \in (2 - c_P, 2]$, with $c_P > 0$ small (see Lemma 5.1). This is eventually reduced to the analogous unweighted improving estimate using some properties of the Cramér approximant, in particular, Corollary 4.4.

The second component is the p -adic estimates, in which the averaging operator in equation (1.5) ends up being replaced by the variant

$$A_{\mathbb{Z}_p^\times}(f, g)(x) := \mathbb{E}_{n \in \mathbb{Z}_p^\times} f(x + n)g(x + P(n)).$$

It is necessary to bound the $L^2(\mathbb{Z}_p) \times L^2(\mathbb{Z}_p) \rightarrow L^q(\mathbb{Z}_p)$ norm of this operator by exactly the constant 1 when $q > 2$ is close to 2 and p is large; losing a multiplicative factor such as $1 + O(1/p)$ would not be acceptable as one needs to multiply these constants over all primes p . Fortunately, the effect of restricting to the invertible elements \mathbb{Z}_p^\times of \mathbb{Z}_p is not too severe and the arguments from [13] can be adapted with only a modest amount of effort to avoid any losses of $O(1/p)$ in the constants.

The most delicate step is to adapt the single-scale estimate in equation (1.2) to the weighted setting. As the Peluse–Prendiville theory is somewhat complicated, our approach is to use the approximation theory from [26] to try to replace the approximant Λ_N with an approximant closer to the constant weight 1. With the theory of the Cramér approximant from [25], it is not too difficult to replace Λ_N by a Cramér approximant $\Lambda_{\text{Cramér},w}$ for a smaller parameter w , with error terms polynomial in w . However, a technical problem then arises: this approximant is not a pure ‘Type I’ sum of the form $\sum_{d|n} \lambda_d$ for certain well-behaved weights λ_d , preventing one from removing the weight entirely. To resolve this, we appeal to the theory from [26] once more to replace the Cramér approximant $\Lambda_{\text{Cramér},w}$ with a more Fourier-analytic approximant, which we call the *Heath-Brown approximant* (as it was introduced by him in [9]). This approximant is defined by

$$\Lambda_{\text{HB},Q}(n) := \sum_{q < Q} \frac{\mu(q)}{\varphi(q)} c_q(n),$$

where Q is a parameter of similar size to w and c_q is a Ramanujan sum; roughly speaking, this approximant is the main term in the Fourier restriction of the von Mangoldt function to major arcs. By using the analysis of the little Gowers uniformity norms of Type I sums from [17], we are able to show that $\Lambda_{\text{Cramér},w}$ is close in these norms to $\Lambda_{\text{HB},w}$ and then, by the theory from [26] (and a dyadic decomposition), one can replace the former by the latter, at least for the purposes of proving an ‘ ℓ^∞ ’ Peluse–Prendiville inverse theorem for weighted averages. As in [13], it is also necessary to obtain a more delicate ‘ ℓ^2 ’ inverse theorem, which requires a weighted version of the ℓ^p improving inequality from [8], but this can be achieved by a variant of the arguments just presented.

Remark 1.5. The proof of Theorem 1.3 quickly yields a version of Peluse’s inverse theorem [23, Theorem 3.3] with prime weights. This was not needed for proving Theorem 1.3 (what we did need was in essence a version with the weight function Λ_N ; see Proposition 5.3), but we believe such a result may be of independent interest, so we record it as Theorem 6.1. Some combinatorial applications of this result will be investigated in a future work.

Remark 1.6. We expect the methods of this paper to be applicable also to pointwise convergence of bilinear polynomial ergodic averages weighted by some other weights of arithmetic interest. The exact requirements for the weight are not so easy to axiomatize, but we need the weight to satisfy analogues of equations (3.1)–(3.4), as well as a suitable ‘local-to-global’ factorization over the primes to be able to pass to the adèles. In particular, we expect the methods to be applicable to ergodic averages weighted by the divisor function τ , but we will not pursue this problem here.

2. Notation

2.1. *General notation.* Our notation largely follows [13], though somewhat abridged, as some of the notation in [13] is only used to establish results or arguments that we are treating here as ‘black boxes’.

We use $\mathbb{Z}_+ := \{1, 2, \dots\}$ to denote the positive integers and $\mathbb{N} := \{0, 1, 2, \dots\}$ to denote the natural numbers.

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, $\mathbb{1}_E(x) = \mathbb{1}_{x \in E}$. We use $|E|$ to denote the cardinality of a set E and adopt for $f: E \rightarrow \mathbb{C}$ the averaging notation

$$\mathbb{E}_{n \in E} f(n) := \frac{1}{|E|} \sum_{n \in E} f(n)$$

if E is finite and non-empty. We similarly define L^p norms

$$\|f\|_{L^p(E)} := \left(\sum_{n \in E} |f(n)|^p \right)^{1/p}$$

for $0 < p < \infty$, with the usual convention that $\|f\|_{L^\infty(E)}$ is the (essential) supremum of f on E . One can extend these averaging conventions to other measurable spaces E of positive finite measure (such as a p -adic group \mathbb{Z}_p equipped with Haar probability measure), if f (or $|f|^p$) is absolutely integrable, in the obvious fashion. When X is equipped with counting measure, we will write $\ell^p(X)$ or just ℓ^p in place of $L^p(X)$.

Throughout, p' denotes the dual exponent of $p \in [1, \infty]$, so $1/p + 1/p' = 1$.

If $f: X \rightarrow \mathbb{C}$, $g: Y \rightarrow \mathbb{C}$ are functions, we use $f \otimes g: X \times Y \rightarrow \mathbb{C}$ to denote the tensor product

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

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 . For any $N \geq 1$, we define the *logarithmic scale* $\text{Log } N$ of N by the formula

$$\text{Log } N := \lfloor \log N / \log 2 \rfloor; \tag{2.1}$$

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

For any two quantities A, B , we will write $A \lesssim B$, $B \gtrsim A$ or $A = O(B)$ to denote the bound $|A| \leq CB$ 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 $|A| \leq C_\rho B$ 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. Due to our reliance in some places on tools based on Siegel's theorem (specifically, Siegel's theorem is used in [18], and we will use results from that paper to establish equation (3.1)), several of the implied constants in our arguments will be ineffective, but we will not track the effectivity of constants explicitly in this paper.

2.3. Algebraic notation. If R is a commutative ring, we use R^\times to denote the multiplicatively invertible elements of R .

2.4. Number theoretic notation. For any $N > 0$, $[N]$ denotes the discrete interval $[N] := \{n \in \mathbb{Z}_+ : n \leq N\}$. If $q_1, q_2 \in \mathbb{Z}_+$, we write $q_1 \mid q_2$ 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 , and $[a, q]$ the least common multiple.

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

In addition to the von Mangoldt function $\Lambda(n)$ and Möbius function $\mu(n)$ already introduced, we will also use the divisor function $\tau(n) := \sum_{d \mid n} 1$ and the Euler totient function $\varphi(n) := |(\mathbb{Z}/n\mathbb{Z})^\times|$.

2.5. Fourier analytic notation. We write $e(\theta) := e^{2\pi i \theta}$ for any real θ , and also $\|\theta\|_{\mathbb{R}/\mathbb{Z}}$ for the distance from θ to the nearest integer.

For a prime p , we let \mathbb{Z}_p be the ring of p -adic integers, defined as the inverse limit of the cyclic groups $\mathbb{Z}/p^j\mathbb{Z}$ for $j \in \mathbb{N}$; this is a compact abelian group equipped with a Haar probability measure. Similarly, let $\hat{\mathbb{Z}}$ be the ring of profinite integers, defined as the inverse limit of the cyclic groups $\mathbb{Z}/Q\mathbb{Z}$ for all positive integers Q ; this is again a compact abelian group with a Haar probability measure, being the direct product of the \mathbb{Z}_p . We use $\mathbb{E}_{\mathbb{Z}_p}$ or $\mathbb{E}_{\hat{\mathbb{Z}}}$ to denote averaging with respect to these compact abelian groups. Finally, we let $\mathbb{A}_{\mathbb{Z}} := \mathbb{R} \times \hat{\mathbb{Z}}$ denote the ring of adelic integers, which is a locally compact abelian group.

We define some Fourier transforms on various locally compact abelian groups.

- (i) Given a summable function $f: \mathbb{Z} \rightarrow \mathbb{C}$, the Fourier transform $\mathcal{F}_{\mathbb{Z}}f: \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{C}$ is defined by the formula

$$\mathcal{F}_{\mathbb{Z}}f(\theta) := \sum_{n \in \mathbb{Z}} f(n)e(-n\theta).$$

- (ii) Given a Schwartz function $f: \mathbb{R} \rightarrow \mathbb{C}$, the Fourier transform $\mathcal{F}_{\mathbb{R}}f: \mathbb{R} \rightarrow \mathbb{C}$ is defined by the formula

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

- (iii) Given a function $f: \hat{\mathbb{Z}} \rightarrow \mathbb{C}$ which is *Schwartz–Bruhat* in the sense that it factors through a function $f_Q: \mathbb{Z}/Q\mathbb{Z} \rightarrow \mathbb{C}$ on a cyclic group, we define the Fourier transform $\mathcal{F}_{\hat{\mathbb{Z}}}f: \mathbb{Q}/\mathbb{Z} \rightarrow \mathbb{C}$ by the formula

$$\mathcal{F}_{\hat{\mathbb{Z}}}f\left(\frac{a}{Q} \bmod 1\right) := \mathbb{E}_{n \in \mathbb{Z}/Q\mathbb{Z}} f_Q(n)e(-an/Q)$$

for any integer a .

- (iv) Given a function $f: \mathbb{A}_{\mathbb{Z}} \rightarrow \mathbb{C}$ which is Schwartz–Bruhat in the sense that it factors through a function $f_Q: \mathbb{R} \times \mathbb{Z}/Q\mathbb{Z} \rightarrow \mathbb{C}$ which is Schwartz in the first variable, we define the Fourier transform $\mathcal{F}_{\mathbb{A}_{\mathbb{Z}}}f: \mathbb{R} \times \mathbb{Q}/\mathbb{Z} \rightarrow \mathbb{C}$ by the formula

$$\mathcal{F}_{\mathbb{A}_{\mathbb{Z}}}f\left(\xi, \frac{a}{Q} \bmod 1\right) := \mathbb{E}_{n \in \mathbb{Z}/Q\mathbb{Z}} \int_{\mathbb{R}} f_Q(x, n)e(-x\xi - an/Q) dx$$

for integer a and $\xi \in \mathbb{R}$, and $\mathcal{F}_{\mathbb{A}_{\mathbb{Z}}}f$ vanishing otherwise.

We refer the reader to [13, §4] for a further discussion of the Fourier transform on such locally compact abelian groups as \mathbb{Z} , \mathbb{R} , \mathbb{Z}_p , $\hat{\mathbb{Z}}$, $\mathbb{Z}/Q\mathbb{Z}$ or $\mathbb{A}_{\mathbb{Z}}$, and the various intertwining relationships among these transforms.

Given a Schwartz symbol $m: \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{C}$, we define the Fourier multiplier T_m on $\ell^2(\mathbb{Z})$ by the formula

$$T_m f(x) := \int_{\mathbb{R}/\mathbb{Z}} m(\xi) \mathcal{F}_{\mathbb{Z}} f(\xi) e(x\xi) d\xi,$$

and, similarly, given a bilinear Schwartz symbol $m: \mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{C}$, define the bilinear Fourier multiplier B_m by the formula

$$B_m(f, g)(x) := \int_{\mathbb{R}/\mathbb{Z}} \int_{\mathbb{R}/\mathbb{Z}} m(\xi, \eta) \mathcal{F}_{\mathbb{Z}} f(\xi) \mathcal{F}_{\mathbb{Z}} g(\eta) e(x(\xi + \eta)) d\xi d\eta.$$

Linear and bilinear multipliers are defined similarly for the other locally compact abelian groups defined here, and obey a certain operator calculus; again, we refer the reader to [13, §4] for details, as we shall largely use facts and arguments about these operators from [13] as ‘black boxes’.

We will need the Ionescu–Wainger Fourier multipliers on major arcs. Again, we shall mostly be using these tools as ‘black boxes’, so their definition and properties are not of critical importance in this paper; however, for sake of completeness, we recall the main definitions from [13]. Given a small parameter ρ , it is possible to assign a *Ionescu–Wainger height* $h(\alpha) = h_{\rho}(\alpha) \in 2^{\mathbb{N}}$ for each $\alpha \in \mathbb{Q}/\mathbb{Z}$; see [13, Appendix A]. Using this height, we define the Ionescu–Wainger arithmetic frequency sets

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

and the Ionescu–Wainger major arcs

$$\mathcal{M}_{\leq l, \leq k} := \{\xi + \alpha : \xi \in \mathbb{R}, |\xi| \leq 2^k, \alpha \in (\mathbb{Q}/\mathbb{Z})_{\leq l}\}; \quad (2.2)$$

thus, $\mathcal{M}_{\leq l, \leq k}$ is the union of arcs $[\alpha - 2^k, \alpha + 2^k]$ for $\alpha \in (\mathbb{Q}/\mathbb{Z})_{\leq l}$; we will be focused on the regime where k is sufficiently small that these arcs are disjoint, which happens whenever $k \leq -C_{\rho} 2^{\rho l}$. We also use the variants

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

and

$$\mathcal{M}_{l, \leq k} := \mathcal{M}_{\leq l, \leq k} \setminus \mathcal{M}_{\leq l-1, \leq k}$$

with the convention that $(\mathbb{Q}/\mathbb{Z})_{\leq -1}$ and $\mathcal{M}_{\leq -1, k}$ are empty.

The *Ionescu–Wainger Fourier projection operator* $\Pi_{\leq l, \leq k}$ for any $(l, k) \in \mathbb{N} \times \mathbb{Z}$ is defined by the formula

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

where η is a smooth even function supported on $[-1, 1]$ that equals 1 on $[-1/2, 1/2]$. We then define

$$\Pi_{l,\leq k} := \Pi_{\leq l,\leq k} - \Pi_{\leq l-1,\leq k}.$$

We refer the reader to [13, §5, Appendix A] for the key properties of these projections, which can be viewed as analogues of Littlewood–Paley projection operators for major arcs.

2.6. Variational norms. A sequence $1 \leq N_1 < \dots < N_k$ of positive reals is said to be λ -lacunary for some $\lambda \geq 1$ if

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

for all $1 \leq j < k$.

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

$$\|(\mathbf{a}_t)_{t \in \mathbb{I}}\|_{V^r(\mathbb{I}; B)} := \sup_{J \in \mathbb{Z}_+} \sup_{\substack{t_0 \leq \dots \leq t_J \\ t_j \in \mathbb{I}}} \left(\sum_{j=0}^{J-1} \|\mathbf{a}(t_{j+1}) - \mathbf{a}(t_j)\|_B^r \right)^{1/r}, \quad (2.3)$$

where the supremum is taken over all finite increasing sequences in \mathbb{I} and is set by convention to equal zero if \mathbb{I} is empty.

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

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

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 $\mathbf{V}^r(\mathbb{I}; X)$ to $\mathbf{V}^r(\mathbb{I})$ or \mathbf{V}^r .

2.7. Gowers norms. In addition to the little Gowers uniformity norm $u^{d+1}[N]$ defined in equation (1.8), we will also need the full Gowers norm $U^{d+1}[N]$ defined for functions $f: \mathbb{Z} \rightarrow \mathbb{C}$ as

$$\|f\|_{U^{d+1}[N]} := \|f\mathbb{1}_{[N]}\|_{U^{d+1}(\mathbb{Z})} / \|\mathbb{1}_{[N]}\|_{U^{d+1}(\mathbb{Z})},$$

where the $U^{d+1}(\mathbb{Z})$ norm is defined for finitely supported functions by the formula

$$\|f\|_{U^{d+1}(\mathbb{Z})}^{2^{d+1}} := \sum_{x, h_1, \dots, h_{d+1} \in \mathbb{Z}} \prod_{\omega \in \{0,1\}^{d+1}} \mathcal{C}^{\omega_1 + \dots + \omega_{d+1}} f\left(x + \sum_{j=1}^{d+1} \omega_j h_j\right),$$

where $\omega = (\omega_1, \dots, \omega_{d+1})$ and \mathcal{C} denotes the complex conjugation operator. It is well known that

$$\|f\|_{u^{d+1}[N]} \lesssim_d \|f\|_{U^{d+1}[N]}; \quad (2.5)$$

see, e.g. [5, equation (2.2)].

Similar uniformity norms $u^{d+1}(I)$, $U^{d+1}(I)$ can then be defined for other intervals $I \subset \mathbb{R}$ than $[N]$ in the obvious fashion.

3. High-level proof of theorem

We now describe the high-level proof of Theorem 1.3, reducing it to two key statements (Theorem 3.2 and Proposition 3.4) that we will prove in §5. The arguments here will closely follow those of [13], and some familiarity with the arguments in that paper would be highly recommended to follow the text in this section.

In the next section, we shall introduce an approximant $\Lambda_N: \mathbb{N} \rightarrow \mathbb{R}$ to Λ (depending on a parameter C_0) which enjoys the bound

$$\|\Lambda - \Lambda_N\|_{u^{d+1}[N]} \lesssim_{A, C_0} \langle \text{Log } N \rangle^{-A} \quad (3.1)$$

for any $A > 0$, as well as the pointwise bound

$$\Lambda_N(n) \lesssim_{C_0} \langle \text{Log } N \rangle^{O(1)}, \quad (3.2)$$

the L^1 bound

$$\mathbb{E}_{n \in [N]} |\Lambda_N(n)| \lesssim_{C_0} 1 \quad (3.3)$$

and finally the *polynomial improving bound*

$$\|\mathbb{E}_{n \in [N]} (\Lambda(n) + |\Lambda_N(n)|) |g(\cdot - P(n) + n)|\|_{\ell^{p'}(\mathbb{Z})} \lesssim_{C_0} N^{d(1/p' - 1/p)} \|g\|_{\ell^p(\mathbb{Z})} \quad (3.4)$$

for all $u_P < p \leq 2$ and $g \in \ell^p(\mathbb{Z})$, with $u_P < 2$ an exponent depending only on P , and $C > 0$ a constant also depending only on P .

We shall also require further properties of Λ_N in the following as needed. (Our choice of approximant Λ_N will in fact be non-negative and, although this is not crucial, it makes it easier to establish the L^1 bound in equation (3.3) and the improving bound in equation (3.4).)

Arguing as in the proof of [13, Proposition 3.2(i)] (inserting the non-negative weight Λ as necessary), we see that the pointwise convergence claim of Theorem 1.3 follows from the ‘Hölder variational estimate’ in equation (1.9), so we focus now on this estimate. Henceforth, we fix $p_1, p_2, p, d, P, r, \lambda$, as well as the finite λ -lacunary set \mathbb{D} . We allow all constants to depend on $p_1, p_2, p, d, P, r, \lambda$ (but not on \mathbb{D}). As in [13, §5], we now select sufficiently large parameters

$$1 \lesssim C_0 \lesssim C_1 \lesssim C_2 \lesssim C_3.$$

By a routine application of Calderón’s transference principle [13, Theorem 3.2(ii)], adapted to this weighted setting, it suffices to prove equation (1.9) for the integer shift system $(\mathbb{Z}, |\cdot|, x \mapsto x - 1)$, endowed with counting measure $|\cdot|$. Thus, our task is now to show that

$$\|(\Lambda_{N, \Lambda; \mathbb{Z}}(f, g))_{N \in \mathbb{D}}\|_{\ell^p(\mathbb{Z}; \mathbf{V}^r)} \lesssim \|f\|_{\ell^{p_1}(\mathbb{Z})} \|g\|_{\ell^{p_2}(\mathbb{Z})}$$

for all $f \in \ell^{p_1}(\mathbb{Z})$ and $g \in \ell^{p_2}(\mathbb{Z})$. Arguing as in the proof of [13, Proposition 3.2(iii)] (inserting the weight Λ as needed), it suffices to prove the ‘upper half’

$$\|(\tilde{\Lambda}_{N, \Lambda}(f, g))_{N \in \mathbb{D}}\|_{\ell^p(\mathbb{Z}; \mathbf{V}^r)} \lesssim \|f\|_{\ell^{p_1}(\mathbb{Z})} \|g\|_{\ell^{p_2}(\mathbb{Z})} \quad (3.5)$$

of this estimate, where the averaging operators $\tilde{\Lambda}_{N, w}$ were defined in equation (1.1).

The next step is to replace the von Mangoldt weight Λ by the approximant Λ_N .

LEMMA 3.1. (From Λ to Λ_N) To prove equation (3.5) (and hence, equation (1.9)), it suffices to show that

$$\|(\tilde{A}_{N,\Lambda_N}(f, g))_{N \in \mathbb{D}}\|_{\ell^p(\mathbb{Z}; \mathbf{V}^r)} \lesssim_{C_3} \|f\|_{\ell^{p_1}(\mathbb{Z})} \|g\|_{\ell^{p_2}(\mathbb{Z})}. \quad (3.6)$$

Proof. Assuming equation (3.6), from the triangle inequality and the lacunarity of \mathbb{D} , we see that equation (3.5) reduces to the single-scale estimate

$$\|\tilde{A}_{N,\Lambda-\Lambda_N}(f, g)\|_{\ell^p(\mathbb{Z})} \lesssim_{C_3} \langle \text{Log } N \rangle^{-2} \|f\|_{\ell^{p_1}(\mathbb{Z})} \|g\|_{\ell^{p_2}(\mathbb{Z})}$$

for each $N \in \mathbb{D}$.

Using the triangle and Hölder inequalities, the prime number theorem and the hypothesis in equation (3.3), we may bound

$$\|\tilde{A}_{N,\Lambda-\Lambda_N}(f, g)\|_{\ell^p(\mathbb{Z})} \lesssim_{C_0} \|f\|_{\ell^{p_1}(\mathbb{Z})} \|g\|_{\ell^{p_2}(\mathbb{Z})},$$

so by interpolation (modifying the exponents p_1, p_2, p as needed), it suffices to prove the $\ell^2 \times \ell^2 \rightarrow \ell^1$ bound

$$\|\tilde{A}_{N,\Lambda-\Lambda_N}(f, g)\|_{\ell^1(\mathbb{Z})} \lesssim_{A,C_3} \langle \text{Log } N \rangle^{-A} \|f\|_{\ell^2(\mathbb{Z})} \|g\|_{\ell^2(\mathbb{Z})} \quad (3.7)$$

for any $A > 0$.

We claim that it suffices to prove equation (3.7) when f, g are supported on intervals of length N^d . Write

$$f = \sum_{i \in \mathbb{Z}} f_i, \quad g = \sum_{i \in \mathbb{Z}} g_i, \quad f_i = f \mathbf{1}_{(iN^d, (i+1)N^d]}, \quad g_i = g \mathbf{1}_{(iN^d, (i+1)N^d]}.$$

Let $C = C_P$ be such that $\{P(n) : n \in [N]\}$ is contained in an interval of length CN^d . Supposing that equation (3.7) holds whenever f, g are supported on intervals of length N^d , by the triangle inequality and Cauchy–Schwarz, we have

$$\begin{aligned} \|\tilde{A}_{N,\Lambda-\Lambda_N}(f, g)\|_{\ell^1(\mathbb{Z})} &\lesssim_{A,C_3} \langle \text{Log } N \rangle^{-A} \sum_{\substack{i,j \in \mathbb{Z} \\ |i-j| \leq C+1}} \|f_i\|_{\ell^2(\mathbb{Z})} \|g_j\|_{\ell^2(\mathbb{Z})} \\ &\lesssim_C \langle \text{Log } N \rangle^{-A} \max_{k \in \mathbb{Z}} \sum_{i \in \mathbb{Z}} \|f_i\|_{\ell^2(\mathbb{Z})} \|g_{i+k}\|_{\ell^2(\mathbb{Z})} \\ &\leq \langle \text{Log } N \rangle^{-A} \max_{k \in \mathbb{Z}} \left(\sum_{i \in \mathbb{Z}} \|f_i\|_{\ell^2(\mathbb{Z})}^2 \right)^{1/2} \left(\sum_{i \in \mathbb{Z}} \|g_{i+k}\|_{\ell^2(\mathbb{Z})}^2 \right)^{1/2} \\ &\leq \langle \text{Log } N \rangle^{-A} \|f\|_{\ell^2(\mathbb{Z})} \|g\|_{\ell^2(\mathbb{Z})}. \end{aligned}$$

Assume henceforth that f, g are supported on intervals of length N^d in equation (3.7). By translation, we can further assume that g is supported on $[N^d]$.

By duality, for some function $h \in \ell^\infty(\mathbb{Z})$ with $|h| \leq 1$, we have

$$\|\tilde{A}_{N,\Lambda-\Lambda_N}(f, g)\|_{\ell^1(\mathbb{Z})} = \left| \sum_{x \in \mathbb{Z}} h(x) \tilde{A}_{N,\Lambda-\Lambda_N}(f, g)(x) \right| = \left| \sum_{x \in \mathbb{Z}} f(x) \tilde{A}_{N,\Lambda-\Lambda_N}^*(h, g)(x) \right|, \quad (3.8)$$

where

$$\tilde{A}_{N,\Lambda-\Lambda_N}^*(h, g)(x) := \mathbb{E}_{n \in [N]} (\Lambda - \Lambda_N)(n) h(x+n) g(x - P(n) + n)$$

is one of the adjoint averaging operators. By Cauchy–Schwarz, the desired estimate in equation (3.7) follows from equation (3.8) if we show that

$$\|\tilde{A}_{N,\Lambda-\Lambda_N}^*(h, g)(x)\|_{\ell^2(\mathbb{Z})} \lesssim_{A,C_3} \langle \text{Log } N \rangle^{-A} \|g\|_{\ell^2(\mathbb{Z})}.$$

By equation (3.4) and the triangle inequality, for all $u_P < q \leq 2$, we have

$$\|\tilde{A}_{N,\Lambda-\Lambda_N}^*(h, g)\|_{\ell^{q'}(\mathbb{Z})} \leq \|\tilde{A}_{N,\Lambda-\Lambda_N}^*(1, |g|)\|_{\ell^{q'}(\mathbb{Z})} \lesssim N^{d(1/q'-1/q)} \|g\|_{\ell^q(\mathbb{Z})}. \quad (3.9)$$

However, [26, Theorem 4.1] (i.e. equation (1.7)), the assumption on the support of g and the hypotheses in equations (3.1) and (3.2), we have

$$\|\tilde{A}_{N,\Lambda-\Lambda_N}^*(h, g)\|_{\ell^1(\mathbb{Z})} \lesssim_{A,C_3} \langle \text{Log } N \rangle^{-A} N^d \|g\|_{\ell^\infty(\mathbb{Z})} \quad (3.10)$$

for any $A > 0$. Interpolating equations (3.9) and (3.10), the claim in equation (3.7) follows. \square

With this lemma, we can now pass to the approximant Λ_N .

We are left with showing equation (3.6). Note from equation (3.3) and the triangle and Hölder inequalities that \tilde{A}_{N,Λ_N} is bounded from $\ell^{p_1}(\mathbb{Z}) \times \ell^{p_2}(\mathbb{Z})$ to $\ell^p(\mathbb{Z})$ whenever $1/p_1 + 1/p_2 = 1/p$; the challenge is to estimate all the scales N in \mathbb{D} simultaneously in \mathbf{V}^r norm. We can restrict attention to scales $N \geq C_3$, since the contribution of the case $N < C_3$ can be handled just from the Hölder and triangle inequalities. The fact that the weight function Λ_N now depends on N will not significantly impact the arguments that follow.

As in [13, §5], we introduce the Ionescu–Wainger parameter

$$\rho := 1/C_1.$$

We use c to denote various small positive constants that can depend on the fixed quantities p_1, p_2, d, P, r , but do not depend on C_0, C_1, C_2, C_3 (or ρ). As reviewed in §2.5, this allows us to create major arc sets $\mathcal{M}_{\leq l, \leq k}$, $\mathcal{M}_{l, \leq k}$ for $l \in \mathbb{N}, k \in \mathbb{Z}$, as well as associated Ionescu–Wainger multipliers $\Pi_{\leq l, \leq k}$, $\Pi_{l, \leq k}$. As in [13, equation (5.8)], we say that the pair (l, k) has *good major arcs* if

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

for some sufficiently large C_ρ depending only on ρ . This condition will always be satisfied in practice and will ensure that the intervals $[\alpha - 2^k, \alpha + 2^k]$ that comprise $\mathcal{M}_{\leq l, \leq k}$ in equation (2.2) are disjoint; thus, avoiding any difficulties arising from ‘aliasing’.

In §5, we shall establish the following crucial variant of [13, Theorem 5.12].

THEOREM 3.2. (Single scale minor arc estimate) *Let $N \geq 1, l \in \mathbb{N}$, and suppose that $f, g \in \ell^2(\mathbb{Z})$ obey one of the following two properties:*

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

Then, one has

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

As in [13, equation (5.22)], we introduce the scales

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

and repeat the arguments in [13, §5] all the way to [13, equation (5.25)], inserting the weight Λ_N as needed, to reduce to establishing the bound

$$\begin{aligned} & \|(\tilde{A}_{N, \Lambda_N}(\Pi_{l_1, \leq -\operatorname{Log} N + l_{(N)}} f, \Pi_{l_2, \leq -d \operatorname{Log} N + dl_{(N)}} g))_{N \in \mathbb{D}: l_1, l_2 \leq l_{(N)}}\|_{\ell^{p_0}(\mathbb{Z}; \mathbb{V}^r)} \\ & \lesssim_{C_3} 2^{-\rho l} \|f\|_{\ell^2(\mathbb{Z})} \|g\|_{\ell^2(\mathbb{Z})} \end{aligned}$$

for all $l_1, l_2 \in \mathbb{N}$, where $l := \max(l_1, l_2)$.

Now, we fix l_1, l_2 , and (as in [13, equation (5.26)]) introduce the quantity

$$u := \lfloor C_2 2^{2\rho l} \rfloor. \quad (3.11)$$

As in [13, equations (5.27), (5.28)], we introduce the frequency-localized functions

$$F_N^{u, l_1, s_1} := \begin{cases} \Pi_{l_1, \leq -\operatorname{Log} N + s_1} f - \Pi_{l_1, \leq -\operatorname{Log} N + s_1 - 1} f, & s_1 > -u, \\ \Pi_{l_1, \leq -\operatorname{Log} N - u} f, & s_1 = -u, \end{cases} \quad (3.12)$$

and

$$G_N^{u, l_2, s_2} := \begin{cases} \Pi_{l_2, \leq d(-\operatorname{Log} N + s_2)} g - \Pi_{l_2, \leq d(-\operatorname{Log} N + s_2 - 1)} g, & s_2 > -u, \\ \Pi_{l_2, \leq d(-\operatorname{Log} N - u)} g, & s_2 = -u, \end{cases} \quad (3.13)$$

for any integers $-u \leq s_1, s_2 \leq l_{(N)}$. Arguing as in the text up to [13, Theorem 5.30], inserting the weight Λ_N as necessary, it now suffices to establish the following.

THEOREM 3.3. (Variational paraproduct estimates) *Let $l_1, l_2 \in \mathbb{N}$, $l := \max(l_1, l_2)$, let $f, g: \mathbb{Z} \rightarrow \mathbb{C}$ be finitely supported and define u by equation (3.11). 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 equations (3.12) and (3.13) and*

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

Then,

$$\begin{aligned} & \|(\tilde{A}_{N, \Lambda_N}(F_N, G_N))_{N \in \mathbb{I}}\|_{\ell^p(\mathbb{Z}; \mathbb{V}^r)} \\ & \lesssim_{C_3} \langle \max(l, s_1, s_2) \rangle^{O(1)} 2^{O(\rho l) - c \max(l, s_1, s_2)} \mathbb{I}_{p_1=p_2=2} \|f\|_{\ell^{p_1}(\mathbb{Z})} \|g\|_{\ell^{p_2}(\mathbb{Z})}. \end{aligned} \quad (3.14)$$

Repeating the proof of [13, Proposition 5.33], inserting the weight Λ_N as needed, we see that Theorem 3.3 already holds in the ‘high-high’ case where $s_1, s_2 > -u$ and $p_1 = p_2 = 2$. Thus, we may assume that at least one of the statements $s_1 = -u, s_2 = -u$ or $(p_1, p_2) \neq (2, 2)$ holds.

We now begin the arguments in [13, §7]. We introduce the functions

$$F := \Pi_{l_1, \leq -u} f, \quad G := \Pi_{l_2, \leq -u} g$$

and note that

$$F_N = T_{\varphi_N}^{l_1} F, \quad G_N = T_{\varphi_N}^{l_2} G,$$

where

$$\varphi_N(\xi) := \begin{cases} \eta(2^{\log N - s_1} \xi) - \eta(2^{\log N - s_1 + 1} \xi), & s_1 > -u, \\ \eta(2^{\log N + u} \xi), & s_1 = -u, \end{cases} \quad (3.15)$$

and

$$\tilde{\varphi}_N(\xi) := \begin{cases} \eta(2^{d(\log N - s_2)} \xi) - \eta(2^{d(\log N - s_2 + 1)} \xi), & s_2 > -u, \\ \eta(2^{d(\log N + u)} \xi), & s_2 = -u. \end{cases} \quad (3.16)$$

Repeating the arguments up to [13, equation (7.7)], we thus see that it suffices to show that the tuple

$$(\tilde{A}_{N, \Lambda_N}(T_{\varphi_N}^{l_1} F, T_{\tilde{\varphi}_N}^{l_2} G))_{N \in \mathbb{I}}$$

is ‘acceptable’ in the sense that it has an $\ell^{p_0}(\mathbb{Z}; \mathbf{V}^r)$ norm of

$$\lesssim_{C_3} (\max(l, s_1, s_2))^{O(1)} 2^{O(\rho l) - c \max(l, s_1, s_2)} \mathbb{1}_{p_1 = p_2 = 2} \|F\|_{\ell^{p_1}(\mathbb{Z})} \|G\|_{\ell^{p_1}(\mathbb{Z})}.$$

We introduce the arithmetic symbol $m_{\hat{\mathbb{Z}}^\times} : (\mathbb{Q}/\mathbb{Z})^2 \rightarrow \mathbb{C}$ by the formula

$$m_{\hat{\mathbb{Z}}^\times} \left(\frac{a_1}{q} \bmod 1, \frac{a_2}{q} \bmod 1 \right) = \mathbb{E}_{n \in (\mathbb{Z}/q\mathbb{Z})^\times} e \left(\frac{a_1 n + a_2 P(n)}{q} \right) \quad (3.17)$$

for any $q \in \mathbb{Z}_+$ and $a_1, a_2 \in \mathbb{Z}$; this differs from the corresponding symbol $m_{\hat{\mathbb{Z}}}$ in [13] by restricting n to the primitive residue classes of $\mathbb{Z}/q\mathbb{Z}$ rather than all residue classes, which is a key effect of weighting by Λ . It is easy to see from the Chinese remainder theorem that $m_{\hat{\mathbb{Z}}^\times}$ is well defined, in the sense that replacing a_1, a_2, q by ka_1, ka_2, kq for any positive integer k does not affect the right-hand side of equation (3.17). Given any Schwartz function $m : \mathbb{R}^2 \rightarrow \mathbb{C}$, we then define the twisted bilinear multiplier operator $B_m^{l_1, l_2, m_{\hat{\mathbb{Z}}^\times}}(f, g)$ for rapidly decreasing $f, g : \mathbb{Z} \rightarrow \mathbb{C}$ by the formula

$$\begin{aligned} B_m^{l_1, l_2, m_{\hat{\mathbb{Z}}^\times}}(f, g)(x) &:= \sum_{\alpha_1 \in (\mathbb{Q}/\mathbb{Z})_{l_1}, \alpha_2 \in (\mathbb{Q}/\mathbb{Z})_{l_2}} m_{\hat{\mathbb{Z}}^\times}(\alpha_1, \alpha_2) \\ &\times \int_{\mathbb{R}^2} m(\xi_1, \xi_2) \mathcal{F}_{\mathbb{Z}} f(\alpha_1 + \xi_1) \mathcal{F}_{\mathbb{Z}} g(\alpha_2 + \xi_2) e(-x(\alpha_1 + \alpha_2 + \xi_1 + \xi_2)) d\xi_1 d\xi_2. \end{aligned}$$

As in [13, equation (7.9)], we also introduce the continuous symbol $\tilde{m}_{N, \mathbb{R}} : \mathbb{R}^2 \rightarrow \mathbb{C}$ by the formula

$$\tilde{m}_{N, \mathbb{R}}(\xi_1, \xi_2) := \int_{1/2}^1 e(\xi_1 N t + \xi_2 P(N t)) dt$$

and also the cutoff functions

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

for any integer k and frequency $\xi \in \mathbb{R}$, where $\eta : \mathbb{R} \rightarrow [0, 1]$ is a fixed smooth even function supported on $[-1, 1]$ that equals one on $[-1/2, 1/2]$.

In §5, we will prove the following analogue of [13, Proposition 7.13].

PROPOSITION 3.4. (Major arc approximation of \tilde{A}_{N,Λ_N}) For any $N \geq 1$ and $s \in \mathbb{N}$ with $-\log N + s \leq -u$, we have

$$\begin{aligned} & \|\tilde{A}_{N,\Lambda_N}(\Pi_{I_1, \leq -\log N+s} \tilde{F}, \Pi_{I_2, \leq -d \log N+ds} \tilde{G}) \\ & \quad - B_{(\eta \leq -\log N+s \otimes \eta \leq -d \log N+ds)}^{I_1, I_2, m_{\hat{\mathbb{Z}} \times}}(\tilde{F}, \tilde{G})\|_{\ell^p(\mathbb{Z})} \\ & \lesssim_{C_3} 2^{O(\max(2^{\rho l}, s))} \exp(-\log^c N) \|\tilde{F}\|_{\ell^{p_1}(\mathbb{Z})} \|\tilde{G}\|_{\ell^{p_2}(\mathbb{Z})} \end{aligned} \quad (3.18)$$

for all $\tilde{F} \in \ell^{p_1}(\mathbb{Z})$, $\tilde{G} \in \ell^{p_2}(\mathbb{Z})$.

This is a slightly weaker type of bound than the corresponding result in [13], as the polynomial gain of N^{-1} has been reduced to the quasipolynomial gain of $\exp(-\log^c N)$. However, this is still good enough to dominate the $2^{O(\max(2^{\rho l}, s))}$ terms, since from [13, equation (7.1)], one has

$$N \geq \max(2^{2^{\max(l, s_1, s_2)/2}}, C_3) \quad (3.19)$$

for all $N \in \mathbb{I}$. Because of this, we can repeat the Fourier-analytic arguments in [13, §7] down to [13, Theorem 7.23] with the obvious changes, and reduce to showing the acceptability of the small-scale model tuple

$$\left(\int_{1/2}^1 B_{m_*}^{I_1, I_2, m_{\hat{\mathbb{Z}} \times}}(T_{\varphi_{N,t}}^{I_1} F, T_{\tilde{\varphi}_{N,t}}^{I_2} G) dt \right)_{N \in \mathbb{I}_{\leq}} \quad (3.20)$$

and the large-scale model tuple

$$\left(\int_{1/2}^1 B_{1 \otimes m_{\hat{\mathbb{Z}} \times}}(T_{\varphi_{N,t} \otimes 1} F_{\mathbb{A}}, T_{\tilde{\varphi}_{N,t} \otimes 1} G_{\mathbb{A}}) \right)_{N \in \mathbb{I}_{>}}, \quad (3.21)$$

where:

- (i) $\mathbb{I}_{\leq} := \{N \in \mathbb{I} : N \leq 2^{2^u}\}$ and $\mathbb{I}_{>} := \{N \in \mathbb{I} : N > 2^{2^u}\}$;
- (ii) $m_*(\xi_1, \xi_2) := \eta_{\leq -2u}(\xi_1) \eta_{\leq -2du}(\xi_2)$;
- (iii) $\varphi_{N,t}(\xi) := \varphi_N(\xi) e(Nt\xi)$, $\tilde{\varphi}_{N,t}(\xi) := \varphi_N(\xi) e(P(Nt)\xi)$;
- (iv) the adelic model functions $F_{\mathbb{A}} \in L^{p_1}(\mathbb{A}_{\mathbb{Z}})$, $G_{\mathbb{A}} \in L^{p_2}(\mathbb{A}_{\mathbb{Z}})$ are defined by the formulae

$$F_{\mathbb{A}}(x, y) := \sum_{\alpha_1 \in (\mathbb{Q}/\mathbb{Z})_{I_1}} \int_{\mathbb{R}} \eta_{\leq -2^{u-1}}(\xi_1) \mathcal{F}_{\mathbb{Z}} F(\alpha_1 + \xi_1) e(-(\xi_1, \alpha_1) \cdot (x, y)) d\xi_1 \quad (3.22)$$

and

$$G_{\mathbb{A}}(x, y) := \sum_{\alpha_2 \in (\mathbb{Q}/\mathbb{Z})_{I_2}} \int_{\mathbb{R}} \eta_{\leq -2^{u-1}}(\xi_2) \mathcal{F}_{\mathbb{Z}} G(\alpha_2 + \xi_2) e(-(\xi_2, \alpha_2) \cdot (x, y)) d\xi_2 \quad (3.23)$$

for $x \in \mathbb{R}$, $y \in \hat{\mathbb{Z}}$.

We can then repeat the integration by parts arguments in the remainder of [13, §7] (replacing $m_{\hat{\mathbb{Z}}}$ by $m_{\hat{\mathbb{Z}}^\times}$) and reduce to establishing the small-scale model estimate

$$\begin{aligned} & \|(\mathbf{B}_{m_*}^{l_1, l_2, m_{\hat{\mathbb{Z}}^\times}}(\mathbf{T}_{\varphi_{N,t,j_1}}^{l_1} F, \mathbf{T}_{\tilde{\varphi}_{N,t,j_2}}^{l_2} G))_{N \in \mathbb{I}_{\leq}}\|_{\ell^p(\mathbb{Z}; \mathbf{V}^r)} \\ & \lesssim_{C_3} \langle \max(l, s_1, s_2) \rangle^{O(1)} 2^{O(\rho l) - cl \mathbb{1}_{p_1=p_2=2}} \|F\|_{\ell^{p_1}(\mathbb{Z})} \|G\|_{\ell^{p_2}(\mathbb{Z})} \end{aligned} \quad (3.24)$$

and the large-scale model estimate

$$\begin{aligned} & \|(\mathbf{B}_{1 \otimes m_{\hat{\mathbb{Z}}^\times}}(\mathbf{T}_{\varphi_{N,t,j_1} \otimes 1} F_{\mathbb{A}}, \mathbf{T}_{\tilde{\varphi}_{N,t,j_2} \otimes 1} G_{\mathbb{A}}))_{N \in \mathbb{I}_{>}}\|_{L^p(\mathbb{A}_{\mathbb{Z}}; \mathbf{V}^r)} \\ & \lesssim_{C_3} \langle \max(l, s_1, s_2) \rangle^{O(1)} 2^{O(\rho l) - cl \mathbb{1}_{p_1=p_2=2}} \|F_{\mathbb{A}}\|_{L^{p_1}(\mathbb{A}_{\mathbb{Z}})} \|G_{\mathbb{A}}\|_{L^{p_2}(\mathbb{A}_{\mathbb{Z}})}, \end{aligned} \quad (3.25)$$

whenever $1/2 \leq t \leq 1$ and $j_1, j_2 \in \{-1, 0, +1\}$ are such that

$$(s_1, j_1), (s_2, j_2) \neq (-u, -1), \quad (3.26)$$

where

$$\varphi_{N,t,j_1}(\xi_1) := (2^{-s_1} N \xi_1)^{j_1} \varphi_{N,t}(\xi_1) \quad (3.27)$$

and

$$\tilde{\varphi}_{N,t,j_2}(\xi_2) := (2^{-ds_2} N^d \xi_2)^{j_2} \tilde{\varphi}_{N,t}(\xi_2). \quad (3.28)$$

To prove the small-scale argument in equation (3.25), we use the two-dimensional Radamacher–Menshov inequality [13, Corollary 8.2] by repeating the arguments of [13, §8] (replacing $m_{\hat{\mathbb{Z}}}$ by $m_{\hat{\mathbb{Z}}^\times}$), reducing matters to establishing the following single-scale estimate.

LEMMA 3.5. (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*

$$\|\mathbf{B}_{m_*}^{l_1, l_2, m_{\hat{\mathbb{Z}}^\times}}(\tilde{F}, \tilde{G})\|_{\ell^p(\mathbb{Z})} \lesssim_{C_3} 2^{-cl \mathbb{1}_{p_1=p_2=2}} \|\tilde{F}\|_{\ell^{p_1}(\mathbb{Z})} \|\tilde{G}\|_{\ell^{p_2}(\mathbb{Z})}.$$

However, this can be proven by repeating the proof of [13, Lemma 8.6], using Proposition 3.4 in place of [13, Proposition 7.13]; the replacement of $m_{\hat{\mathbb{Z}}}$ with $m_{\hat{\mathbb{Z}}^\times}$ makes no difference here, and the slight reduction in strength of Proposition 3.4 from a polynomial gain in N to a quasipolynomial gain in N is similarly manageable.

It remains to establish the large-scale estimate in equation (3.25). We repeat the arguments in [13, §9], replacing $m_{\hat{\mathbb{Z}}}$ by $m_{\hat{\mathbb{Z}}^\times}$, and noting that $\mathbf{B}_{1 \otimes m_{\hat{\mathbb{Z}}^\times}}$ is the tensor product of the identity and the bilinear operator $A_{\hat{\mathbb{Z}}^\times}$ on the profinite integers defined for $f: \mathbb{Z}/Q\mathbb{Z} \rightarrow \mathbb{C}$, $g: \mathbb{Z}/Q\mathbb{Z} \rightarrow \mathbb{C}$ for any Q (which one can also view as functions on $\hat{\mathbb{Z}}$ in the obvious fashion) by the formula

$$A_{\hat{\mathbb{Z}}^\times}(f, g)(x) := \mathbb{E}_{n \in (\mathbb{Z}/Q\mathbb{Z})^\times} f(x+n)g(x+P(n)).$$

These arguments reduce matters to establishing the following analogue of [13, Theorem 9.9].

THEOREM 3.6. (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 $1 \leq r < (2d/(d-1))$, one has

$$\|A_{\hat{\mathbb{Z}}^\times}(f, g)\|_{L^r(\hat{\mathbb{Z}})} \lesssim_{C_{3,r}} 2^{-cr^l} \|f\|_{L^2(\hat{\mathbb{Z}})} \|g\|_{L^2(\hat{\mathbb{Z}})}.$$

Repeating the arguments in [13, §10] up to [13, equations (10.3), (10.4)], using $A_{\hat{\mathbb{Z}}^\times}$ in place of $A_{\hat{\mathbb{Z}}}$ and Theorem 3.2 in place of [13, Theorem 5.12], we see that it suffices to establish the p -adic bound

$$\|A_{\mathbb{Z}_p^\times}\|_{L^2(\mathbb{Z}_p) \times L^2(\mathbb{Z}_p) \rightarrow L^q(\mathbb{Z}_p)} \lesssim_q 1 \quad (3.29)$$

for all primes p , together with the improvement

$$\|A_{\mathbb{Z}_p^\times}\|_{L^2(\mathbb{Z}_p) \times L^2(\mathbb{Z}_p) \rightarrow L^q(\mathbb{Z}_p)} \leq 1 \quad (3.30)$$

whenever $1 \leq q < (2d/(d-1))$ and p is sufficiently large depending on q , where the averaging operator $A_{\mathbb{Z}_p^\times}$ is defined as

$$A_{\mathbb{Z}_p^\times}(f, g)(x) := \mathbb{E}_{n \in \mathbb{Z}_p^\times} f(x+n)g(x+P(n)).$$

Because \mathbb{Z}_p^\times has density $(p-1)/p$ in \mathbb{Z}_p , we have the pointwise bound

$$|A_{\mathbb{Z}_p^\times}(f, g)(x)| \leq \frac{p}{p-1} A_{\mathbb{Z}_p}(|f|, |g|)(x) \quad (3.31)$$

from the triangle inequality, where

$$A_{\mathbb{Z}_p}(f, g)(x) := \mathbb{E}_{n \in \mathbb{Z}_p} f(x+n)g(x+P(n)).$$

Hence, equation (3.29) is immediate from [13, equation (10.3)]. It remains to establish equation (3.30). As in [13, §10], we may assume $2 < q < (2d/(d-1))$ and $\|f\|_{L^2(\mathbb{Z}_p)} = \|g\|_{L^2(\mathbb{Z}_p)} = 1$ with f, g non-negative, in which case, our task is to show that

$$\mathbb{E}_{n \in \mathbb{Z}_p} |A_{\mathbb{Z}_p^\times}(f, g)(x)|^q \leq 1.$$

Applying equation (3.31) and the bound $\|A_{\mathbb{Z}_p}(|f|, |g|)\|_{L^q(\mathbb{Z}_p)} \leq 1$ from [13, §10] would cost a factor of $(p/(p-1))^q$, which is not acceptable here (the product $\prod_p (p/(p-1))$ diverges). Instead, we follow the arguments in [13, §10], decomposing $f = a + f_0$, $g = b + g_0$, where $0 \leq a, b \leq 1$, f_0, g_0 have mean zero, and the ‘energies’

$$E_f := \|f_0\|_{L^2(\mathbb{Z}_p)}^2, \quad E_g := \|g_0\|_{L^2(\mathbb{Z}_p)}^2$$

obey $0 \leq E_f, E_g \leq 1$ and

$$|a| = (1 - E_f)^{1/2}, \quad |b| = (1 - E_g)^{1/2}.$$

In the case of $A_{\mathbb{Z}_p}$, we clearly have

$$A_{\mathbb{Z}_p}(a, b) = ab, \quad A_{\mathbb{Z}_p}(f_0, b) = 0$$

(was observed in [13, §10]) so that by linearity, we have

$$A_{\mathbb{Z}_p}(f, g) = ab + A_{\mathbb{Z}_p}(f, g_0).$$

For the averaging operator $A_{\mathbb{Z}_p}^\times$, the situation is slightly more complicated; we have

$$A_{\mathbb{Z}_p}^\times(a, b) = ab, \quad A_{\mathbb{Z}_p}^\times(f_0, b) = -\frac{p}{p-1}bh,$$

where $h: \mathbb{Z}_p \rightarrow \mathbb{R}$ is the function

$$h(x) := \mathbb{E}_{n \in \mathbb{Z}_p} f_0(x+n) \mathbb{1}_{p|n}.$$

Since f_0 has mean zero, h has mean zero as well. Furthermore, from Young's convolution inequality, one has the bounds

$$\begin{aligned} \|h\|_{L^2(\mathbb{Z}_p)} &\leq \|f_0\|_{L^2(\mathbb{Z}_p)} \|\mathbb{1}_{p|n}\|_{L^1(\mathbb{Z}_p)} = p^{-1} E_f^{1/2}, \\ \|h\|_{L^q(\mathbb{Z}_p)} &\leq \|f_0\|_{L^2(\mathbb{Z}_p)} \|\mathbb{1}_{p|n}\|_{L^r(\mathbb{Z}_p)} = p^{-1/2-1/q} E_f^{1/2}, \end{aligned} \quad (3.32)$$

where $1/q + 1 = 1/2 + 1/r$.

We now have the decomposition

$$A_{\mathbb{Z}_p}^\times(f, g) = ab + A_{\mathbb{Z}_p}^\times(f, g_0) - \frac{p}{p-1}bh$$

and hence by the Taylor expansion $(x+y)^q = x^q + qx^{q-1}y + O(q^2x^{q-2}y^2)$ (as in [13, §10]), we have

$$\begin{aligned} |A_{\mathbb{Z}_p}^\times(f, g)|^q &= |ab|^q + q|ab|^{q-1} \left(A_{\mathbb{Z}_p}^\times(f, g_0) - \frac{p}{p-1}bh \right) \\ &\quad + O_q(|A_{\mathbb{Z}_p}^\times(f, g_0)|^2 + |A_{\mathbb{Z}_p}^\times(f, g_0)|^q + |h|^2 + |h|^q). \end{aligned}$$

Since $a, b \in [0, 1]$, we can bound $|ab|^q \leq |ab|^2 = (1 - E_f)(1 - E_g)$. Furthermore, $p/(p-1)bh$ has mean zero and $A_{\mathbb{Z}_p}^\times(f, g_0)$ has a mean of at most $\|A_{\mathbb{Z}_p}^\times(f_0, g_0)\|_{L^1(\mathbb{Z}_p)}$ since $A_{\mathbb{Z}_p}^\times(a, g_0)$ has mean zero. We conclude that

$$\begin{aligned} \|A_{\mathbb{Z}_p}^\times(f, g)\|_{L^q(\mathbb{Z}_p)}^q &\leq (1 - E_f)(1 - E_g) + O_q(\|A_{\mathbb{Z}_p}^\times(f_0, g_0)\|_{L^1(\mathbb{Z}_p)}) \\ &\quad + \|A_{\mathbb{Z}_p}^\times(f, g_0)\|_{L^2(\mathbb{Z}_p)}^2 + \|A_{\mathbb{Z}_p}^\times(f, g_0)\|_{L^q(\mathbb{Z}_p)}^q + p^{-2}E_f + p^{-q/2-1}E_f^{q/2}. \end{aligned}$$

By arguing as in [13, §10] (using Theorem 3.2 in place of [13, Theorem 5.12]), we see that if l is any large integer and p is sufficiently large depending on q , we have the estimates

$$\begin{aligned} \|A_{\mathbb{Z}_p}^\times(f_0, g_0)\|_{L^1(\mathbb{Z}_p)} &\lesssim 2^{-c_q l} E_f^{1/2} E_g^{1/2}, \\ \|A_{\mathbb{Z}_p}^\times(f, g_0)\|_{L^2(\mathbb{Z}_p)}^2 &\lesssim 2^{-c_q l} E_g, \\ \|A_{\mathbb{Z}_p}^\times(f, g_0)\|_{L^q(\mathbb{Z}_p)}^q &\lesssim 2^{-c_q l} E_g^{q/2} \end{aligned}$$

for some $c_q > 0$ depending only on q , and hence, by the arithmetic mean-geometric mean inequality and the hypothesis $q > 2$, we have

$$\begin{aligned}\|A_{\mathbb{Z}_p^\times}(f, g)\|_{L^q(\mathbb{Z}_p)}^q &\leq (1 - E_f)(1 - E_g) + O_q((2^{-cq^l} + p^{-2})(E_f + E_g)) \\ &\leq (1 - E_f)(1 - E_g) + O_q((2^{-cq^l} + p^{-2})),\end{aligned}$$

and the right-hand side is bounded by 1 for l and p large enough, as required.

To summarize, to complete the proof of Theorem 1.3, we need to select an approximant Λ_N to the weight Λ at each scale N that obeys the estimates in equations (3.1), (3.2), (3.3) and (3.4), as well as the single scale minor arc estimate in Theorem 3.2 and the major arc approximation in Proposition 3.4. This will be the focus of the next sections.

4. Approximants to the von Mangoldt function

As seen in the previous section, the arguments rely on using an approximant Λ_N to the von Mangoldt function Λ at scale N . There are several plausible candidates for such approximants, including the following.

- (i) Λ itself.
- (ii) A Cramér (or Cramér–Granville) approximant

$$\Lambda_{\text{Cramér}, w}(n) := \frac{W}{\varphi(W)} \mathbb{1}_{(n, W)=1},$$

where

$$W := \prod_{p \leq w} p$$

and $w \geq 1$ is a parameter.

- (iii) A Heath-Brown approximant

$$\Lambda_{\text{HB}, Q}(n) := \sum_{q < Q} \frac{\mu(q)}{\varphi(q)} c_q(n), \quad (4.1)$$

where $Q \geq 1$ is a parameter and $c_q(n)$ are the Ramanujan sums

$$c_q(n) := \sum_{r \in (\mathbb{Z}/q\mathbb{Z})^\times} e(-rn/q). \quad (4.2)$$

Other possibilities for approximants exist, including Goldston–Pintz–Yıldırım type approximants $(\log R) \sum_{\ell|n} \mu(\ell) \eta(\log \ell / \log R)$ and $(\log R) (\sum_{\ell|n} \mu(\ell) \eta(\log \ell / \log R))^2$ for suitable level parameters R and smooth cutoffs η , Selberg sieve approximants $(\sum_{\ell|n} \lambda_\ell)^2$, or adjustments to several of the previous approximants by a correction term arising from a Siegel zero, but we will not discuss these other options further here.

The choice of option (i) (that is, setting $\Lambda_N := \Lambda$) is tempting, particularly in view of recent advances in quantitative understanding of functions such as Λ in [15, 25]. However, it turns out that the presence of a Siegel zero would distort the asymptotics of Λ to such an extent that the desired approximation in Proposition 3.4 no longer holds with quasipolynomial error terms in N , which turns out to significantly complicate the analysis (particularly in the small-scale regime, in which one has to modify the Radamacher–Menshov type arguments significantly). See §6 for further discussion.

The choice of option (ii) has the advantage of being non-negative, reasonably well controlled in ℓ^∞ and also relatively easy to control in Gowers uniformity norms, and so we shall take such a choice for our approximant Λ_N ; specifically, we will set

$$\Lambda_N = \Lambda_{\text{Cramér}, \exp(\text{Log}^{1/C_0} N)}. \quad (4.3)$$

However, there is one aspect in which this approximant $\Lambda_N(n)$ is not ideal: it is not exactly equal to a ‘Type I sum’ $\sum_{\ell|n} \lambda_\ell$, where λ_ℓ are weights supported on relatively small values of d . The Heath-Brown approximants $\Lambda_{\text{HB}, Q}$ introduced in option (iii) are precisely Type I sums, and so we will switch to those approximants at a certain point in the proof.

To achieve these goals, we will need to collect some basic facts about the Cramér approximants $\Lambda_{\text{Cramér}, w}$ and the Heath-Brown approximants $\Lambda_{\text{HB}, Q}$, which may be of independent interest.

4.1. Bounds on the Cramér approximant. We begin with the Cramér approximant. First, we record an easy uniform bound.

LEMMA 4.1. (Uniform bound on Cramér model) *If $w \geq 1$, then*

$$0 \leq \Lambda_{\text{Cramér}, w}(n) \lesssim \langle \text{Log } w \rangle$$

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

Proof. This is immediate from the Mertens theorem bound

$$\frac{W}{\varphi(W)} = \prod_{p \leq w} \frac{p}{p-1} \lesssim \langle \text{Log } w \rangle. \quad \square$$

The Cramér approximant is not easily expressible as an exact Type I sum once w is reasonably large (in particular, larger than $\text{Log } N$), but thanks to the fundamental lemma of sieve theory, it can be approximated by such a sum.

LEMMA 4.2. (Fundamental lemma of sieve theory) *If $2 \leq w \leq y \leq N^{1/10}$, then there exist weights $\lambda_\ell^\pm \in [-1, 1]$, supported on $1 \leq \ell \leq y$, such that*

$$\sum_{\ell|n} \lambda_\ell^- \leq \frac{\varphi(W)}{W} \Lambda_{\text{Cramér}, w}(n) \leq \sum_{\ell|n} \lambda_\ell^+$$

for all n , and also

$$\mathbb{E}_{n \in I} \sum_{\ell|n} \lambda_\ell^\pm = \frac{\varphi(W)}{W} (1 + O(\exp(-\log y / \log w)))$$

for any interval I of length N . In particular,

$$\mathbb{E}_{n \in I} \left| \Lambda_{\text{Cramér}, w}(n) - \frac{W}{\varphi(W)} \sum_{\ell|n} \lambda_\ell^\pm \right| \lesssim \exp(-\log y / \log w).$$

Proof. This follows easily from [11, Lemma 6.3]. □

The fundamental lemma can then be used to give many good estimates for the Cramér model.

PROPOSITION 4.3. (Linear equations in the Cramér model) *Let $t, m \geq 1$ be integers and let $N \geq 100$. Let $\Omega \subset [-N, N]^d$ be convex, and let $\psi_1, \dots, \psi_t: \mathbb{Z}^m \rightarrow \mathbb{Z}$ be linear forms*

$$\psi_i(\vec{n}) = \vec{n} \cdot \vec{\psi}_i + \psi_i(0)$$

for some $\vec{\psi}_i \in \mathbb{Z}^m$ and $\psi_i(0) \in \mathbb{Z}$. Assume that the linear coefficients $\vec{\psi}_1, \dots, \vec{\psi}_t \in \mathbb{Z}^m$ are all pairwise linearly independent and have magnitude at most $\exp(\log^{3/5} N)$. Suppose that $1 \leq z_i \leq \exp(\log^{1/10} N)$ for all $i = 1, \dots, t$. Then, one has

$$\sum_{\vec{n} \in \Omega \cap \mathbb{Z}^m} \prod_{i=1}^t \Lambda_{\text{Cramér}, z_i}(\psi_i(\vec{n})) = \text{vol}(\Omega) \prod_p \beta_p + O_{t,m}(N^m \exp(-c \log^{4/5} N))$$

for some $c > 0$ depending only on t, m , where for each p , β_p is the local factor

$$\beta_p := \mathbb{E}_{\vec{n} \in (\mathbb{Z}/p\mathbb{Z})^m} \prod_{\substack{1 \leq i \leq t \\ p \leq z_i}} \frac{p}{p-1} \mathbb{1}_{\psi_i(\vec{n}) \neq 0},$$

where ψ_i is also viewed as a map from $(\mathbb{Z}/p\mathbb{Z})^m$ to $\mathbb{Z}/p\mathbb{Z}$ in the obvious fashion. Furthermore, β_p obeys the bounds

$$\beta_p = 1 + O_{t,m}(1/p^2) \tag{4.4}$$

for all primes p (and $\beta_p = 1$ if $p > \max(z_1, \dots, z_t)$).

Proof. This is essentially [25, Proposition 5.2] (which relies to a large extent on the fundamental lemma of sieve theory). Strictly speaking, this proposition only covered the case where the z_i were equal to a single parameter z which was also assumed to be at least 2, but an inspection of the argument shows that it applies without significant difficulty to variable z_i as well, even if some of the z_i are as small as 1. The bound in equation (4.4) follows from [25, equations (5.2), (5.5)] (a slightly weaker bound, which also suffices for our application, can be found in [6, Lemma 1.3]). \square

Specializing to the $t = m = 1$ case (and noting that the constant coefficients of ψ_i can be large in Proposition 4.3), we immediately obtain the following corollary.

COROLLARY 4.4. (Mean value of Cramér) *Let $N \geq 100$ and $1 \leq z \leq \exp(\log^{1/10} N)$, then*

$$\mathbb{E}_{n \in I} \Lambda_{\text{Cramér}, z}(n) = 1 + O(\exp(-c \log^{4/5} N))$$

for any interval I of length N . In particular, since $\Lambda_{\text{Cramér}, z}(n)$ is non-negative, we also have

$$\mathbb{E}_{n \in I} |\Lambda_{\text{Cramér}, z}(n)| = 1 + O(\exp(-c \log^{4/5} N)).$$

More generally, if $1 \leq q \leq z$ and $a(q)$ is a residue class, then

$$\mathbb{E}_{n \in I} \Lambda_{\text{Cramér}, z}(n) \mathbb{1}_{n \equiv a(q)} = \frac{\mathbb{1}_{(a, q) = 1}}{\varphi(q)} + O(\exp(-c \log^{4/5} N)).$$

As a more sophisticated application of Proposition 4.3, we record the following improvement of [25, Proposition 1.2].

LEMMA 4.5. (Improved stability of the Cramér model) *If $1 \leq z, w \leq \exp(\text{Log}^{1/10} N)$, for any $d \geq 1$, one has*

$$\|\Lambda_{\text{Cramér},w} - \Lambda_{\text{Cramér},z}\|_{U^{d+1}(I)} \lesssim_d w^{-c} + z^{-c}$$

for any interval I of length N . In particular, by equation (2.5),

$$\|\Lambda_{\text{Cramér},w} - \Lambda_{\text{Cramér},z}\|_{u^{d+1}(I)} \lesssim_d w^{-c} + z^{-c}.$$

In fact, one can take $c = 1/2^{d+1}$ in these estimates.

The result in [25, Proposition 1.2] had an additional term of $\text{Log}^{-c} N$ on the right-hand side. The removal of this term was already conjectured in [25, Remark 5.4].

Proof. Without loss of generality, we may assume that $z \leq w$. Expanding out the expression $\|\Lambda_{\text{Cramér},w} - \Lambda_{\text{Cramér},z}\|_{U^{d+1}(I)}^{2^{d+1}}$ into an alternating sum of 2^{d+1} terms, it suffices to show that

$$\sum_{\epsilon \in \{0,1\}^{d+1}} \sum_{n, h_1, \dots, h_{d+1} \in \mathbb{Z}} \prod_{j=1}^{d+1} \Lambda_{\text{Cramér}, w_\epsilon} \mathbb{1}_I(n + \epsilon_1 h_1 + \dots + \epsilon_{d+1} h_{d+1}) = (X + O(z^{-1})) N^{d+2}$$

for all choices of parameters $w_\epsilon \in \{w, z\}$, where $\epsilon = (\epsilon_1, \dots, \epsilon_{d+1})$ and X is a quantity that is independent of the choice of parameters w_ϵ . Applying Proposition 4.3, the left-hand side is

$$\text{vol}(\Omega) \prod_p \beta_p + O_d(N^{d+2} \exp(-c \text{Log}^{4/5} N)),$$

where Ω is a certain explicit convex polytope of volume $\beta_\infty N^{d+2}$ for some constant β_∞ depending only on d , and the local factors β_p are defined by the formula

$$\beta_p := \mathbb{E}_{n, h_1, \dots, h_{d+1} \in \mathbb{Z}/p\mathbb{Z}} \prod_{\substack{\epsilon \in \{0,1\}^{d+1} \\ p \leq w_\epsilon}} \frac{p}{p-1} \mathbb{1}_{p \nmid n + \epsilon_1 h_1 + \dots + \epsilon_{d+1} h_{d+1}}.$$

The local factors β_p are independent of the w_ϵ if $p \leq w$ or $p > z$. Thus, by equation (4.4), the product $\prod_p \beta_p$ can be written as $Y(1 + O(1/z))$ for some Y that is independent of the w_ϵ parameters, and the claim follows. \square

4.2. *Bounds on the Heath-Brown approximant.* We now turn to the Heath-Brown approximants $\Lambda_{\text{HB}, Q}$. The nice bounds in ℓ^∞ or ℓ^1 one has in Lemma 4.1 or Corollary 4.4 are unfortunately not available for this approximant. However, we have reasonable control in other norms such as ℓ^2 , in large part due to a good Type I representation.

LEMMA 4.6. (Moment bounds for Heath-Brown approximant) *For any $Q \geq 1$, one has the Type I representation*

$$\Lambda_{\text{HB},Q}(n) = \sum_{\substack{\ell|n \\ \ell < Q}} \lambda_\ell \quad (4.5)$$

for some weights λ_ℓ with

$$\lambda_\ell \lesssim \langle \text{Log } Q \rangle. \quad (4.6)$$

In particular, we have the pointwise bound

$$\Lambda_Q(n) \lesssim \tau(n, Q) \langle \text{Log } Q \rangle, \quad (4.7)$$

where $\tau(n, Q)$ is the truncated divisor function

$$\tau(n, Q) := \sum_{\substack{\ell|n \\ \ell < Q}} 1.$$

Furthermore, we have the moment bounds

$$\mathbb{E}_{n \in [N]} |\Lambda_Q(n)|^k \lesssim_k \langle \text{Log } Q \rangle^{2^k + k} \quad (4.8)$$

for any positive integer k and $N \geq 1$.

Proof. Applying the standard identity $c_q(n) = \sum_{\ell|(q,n)} \ell \mu(q/\ell)$ and then writing $q = \ell r$, we have

$$\begin{aligned} \Lambda_Q(n) &= \sum_{q < Q} \frac{\mu(q)}{\varphi(q)} \sum_{\ell|(q,n)} \ell \mu(q/\ell) \\ &= \sum_{\substack{\ell|n \\ \ell < Q}} \frac{\mu(\ell)\ell}{\varphi(\ell)} \sum_{\substack{r < Q/\ell \\ (\ell,r)=1}} \frac{\mu^2(r)}{\varphi(r)}. \end{aligned}$$

We then take

$$\lambda_\ell := \frac{\mu(\ell)\ell}{\varphi(\ell)} \sum_{\substack{r < Q/\ell \\ (\ell,r)=1}} \frac{\mu^2(r)}{\varphi(r)}.$$

From Rankin's trick and Mertens's theorem, for any $1 \leq d \leq Q$, one has

$$\begin{aligned} \sum_{\substack{r \leq Q/\ell \\ (d,\ell)=1}} \frac{\mu^2(r)}{\varphi(r)} &\lesssim \sum_{\substack{r \geq 1 \\ (\ell,r)=1}} \frac{\mu^2(r)}{\varphi(r)r^{1/(\text{Log } Q)}} \\ &\lesssim \prod_{\substack{p \\ p \nmid \ell}} \left(1 + \frac{1}{(p-1)p^{1/(\text{Log } Q)}} \right) \\ &\lesssim \frac{\varphi(\ell)}{\ell} \prod_p \left(1 + \frac{1}{p^{1+1/(\text{Log } Q)}} + o\left(\frac{1}{p^2}\right) \right) \\ &\lesssim \frac{\varphi(\ell)}{\ell} \langle \text{Log } Q \rangle, \end{aligned}$$

where we used the Euler product formula and the standard bound $\zeta(\sigma) \sim 1/(\sigma - 1)$ for $\sigma > 1$ to estimate the product over the primes. This gives equation (4.6). The bound in equation (4.7) then follows from the triangle inequality.

Now, we turn to equation (4.8). We may assume that $Q \geq 100$, as the claim is trivial otherwise. We allow all implied constants to depend on k . In view of equation (4.7), it suffices to establish the bound

$$\sum_{n \in [N]} \tau(n, Q)^k \lesssim N (\text{Log } Q)^{2^k}.$$

We expand

$$\sum_{n \in [N]} \tau(n, Q)^k = \sum_{n \in [N]} \left(\sum_{\substack{\ell | n \\ \ell < Q}} 1 \right)^k = \sum_{n \in [N]} \sum_{\ell_1, \dots, \ell_k < Q} 1 = \sum_{\ell_1, \dots, \ell_k < Q} \frac{N}{[\ell_1, \dots, \ell_k]},$$

where $[a_1, \dots, a_k]$ is the least common multiple of a_1, \dots, a_k .

Now, we apply Rankin's trick. For $\ell_i < Q$, we have $\ell_i^{1/(\text{Log } Q)} = O(1)$, and thus,

$$\mathbb{E}_{n \in [N]} \tau(n, Q)^k \lesssim \sum_{\ell_1, \dots, \ell_k} \frac{1}{\ell_1^{1/\log Q} \dots \ell_k^{1/(\text{Log } Q)} [\ell_1, \dots, \ell_k]}.$$

Factorizing into an Euler product, we conclude that

$$\mathbb{E}_{n \in [N]} \tau(n, Q)^k \lesssim \prod_p \left(1 + \sum_{\substack{a_1, \dots, a_k \in \{0, 1\} \\ (a_1, \dots, a_k) \neq \mathbf{0}}} \frac{1}{p^{1+(a_1+\dots+a_k)/(\text{Log } Q)}} + o\left(\frac{1}{p^2}\right) \right),$$

where $\mathbf{0} := (0, \dots, 0)$. Hence, on taking logarithms, it will suffice to show that

$$\sum_p \sum_{\substack{a_1, \dots, a_k \in \{0, 1\} \\ (a_1, \dots, a_k) \neq \mathbf{0}}} p^{-1-(a_1+\dots+a_k)/(\text{Log } Q)} \leq 2^k \log \log Q + O(1).$$

From partial summation and the prime number theorem, we have

$$\begin{aligned} & \sum_{\substack{a_1, \dots, a_k \in \{0, 1\} \\ (a_1, \dots, a_k) \neq \mathbf{0}}} \sum_{p \geq Q} p^{-1-(a_1+\dots+a_k)/(\log Q)} \\ & \leq \sum_{\substack{a_1, \dots, a_k \in \{0, 1\} \\ (a_1, \dots, a_k) \neq \mathbf{0}}} \int_Q^\infty \frac{t^{-1-(a_1+\dots+a_k)/(\text{Log } Q)}}{\log t} dt + O(1) \\ & \leq 2^k \cdot \int_Q^\infty \frac{t^{-1/(\log Q)}}{t \log t} dt + O(1) \lesssim 2^k + O(1). \end{aligned}$$

Moreover, we can use Mertens's theorem to estimate

$$\sum_{\substack{a_1, \dots, a_k \in \{0, 1\} \\ (a_1, \dots, a_k) \neq \mathbf{0}}} \sum_{p < Q} p^{-1-(a_1+\dots+a_k)/(\text{Log } Q)} \leq 2^k \log(\text{Log } Q) + O(1).$$

Combining these bounds gives the result. □

4.3. *Comparing the Cramér and Heath-Brown approximants.* We have a useful comparison theorem between the Cramér and Heath-Brown approximants.

PROPOSITION 4.7. (Comparison between Cramér and Heath-Brown) *Let $N \geq 1$ and $1 \leq w, Q \leq \exp(\text{Log}^{1/20} N)$, and let $d \geq 1$ be an integer. Then,*

$$\|\Lambda_{\text{Cramér},w} - \Lambda_{\text{HB},Q}\|_{u^{d+1}(I)} \lesssim_d w^{-c} + Q^{-c}$$

for any interval I of length N . As a consequence, from Lemma 4.5 and the triangle inequality, we also have

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

whenever $1 \leq Q_1, Q_2 \leq \exp(\text{Log}^{1/20} N)$.

Proof. We allow all implied constants to depend on d . In view of Lemma 4.5 and the triangle inequality, it suffices to establish the bound

$$\|\Lambda_{\text{Cramér},Q} - \Lambda_{\text{HB},Q}\|_{u^{d+1}(I)} \lesssim Q^{-c}$$

for any interval I of length N , that is to say, it suffices to show that

$$|\mathbb{E}_{n \in I}(\Lambda_{\text{Cramér},Q}(n) - \Lambda_{\text{HB},Q}(n))e(R(n))| \lesssim Q^{-c}$$

for any polynomial $R(n) = \sum_{j=0}^d \alpha_j (n - n_I)^d$ of degree at most d with some real coefficients α_j , where n_I denotes the midpoint of I . By subdividing I into smaller intervals and using the triangle inequality (adjusting the coefficients α_j as necessary), we may assume without loss of generality that

$$N \sim \exp(\text{Log}^{20} Q).$$

We can then also assume that Q (and hence N) are large, as the claim is trivial otherwise. In particular, $\text{Log } N = \text{Log}^{O(1)} Q$, which in practice will permit us to absorb all logarithmic factors of N in the analysis below.

Fix the polynomial R . We may of course assume without loss of generality that

$$|\mathbb{E}_{n \in I}(\Lambda_{\text{Cramér},Q}(n) - \Lambda_{\text{HB},Q}(n))e(R(n))| \geq Q^{-1}.$$

Applying Lemma 4.2 (with $w = Q$ and $y = \exp(\text{Log}^{1/10} N)$) as well as Lemma 4.6, we thus have

$$\left| \mathbb{E}_{n \in I} \left(\sum_{\substack{\ell \leq \exp(\text{Log}^{1/10} N) \\ \ell | n}} \lambda_\ell \right) e(R(n)) \right| \geq Q^{-1}$$

for some weights λ_ℓ of size $O(\text{Log}^{O(1)} N) = O(\text{Log}^{O(1)} Q)$. Applying [17, Proposition 2.1] (after shifting the summation variable by n_I), we conclude that the polynomial R is major arc in the sense that there exists an integer $1 \leq q \lesssim Q^{O(1)}$ such that

$$\|q\alpha_j\|_{\mathbb{R}/\mathbb{Z}} \lesssim Q^{O(1)}/N^j$$

for all $1 \leq j \leq d$. We may assume that $q \geq Q$ by multiplying q by an integer of size Q if necessary. Thus, one can write $R(n) = R_0(n) + E(n)$, where R_0 is a polynomial of

degree at most d that is periodic with period q and the error E satisfies $\sup_{n \in I} |E(n+1) - E(n)| = O(Q^{O(1)}/N)$.

Set

$$w := q, \quad W := \prod_{p < w} p,$$

and thus, $Q \leq w \lesssim Q^{O(1)}$. By Lemma 4.5 and the triangle inequality, it will suffice to show that

$$|\mathbb{E}_{n \in I} (\Lambda_{\text{Cramér}, w}(n) - \Lambda_{\text{HB}, Q}(n)) e(R(n))| \lesssim Q^{-c}.$$

Breaking up I into intervals J of length \sqrt{N} and using the slowly varying nature of $E(n)$, it suffices to show that

$$|\mathbb{E}_{n \in J} (\Lambda_{\text{Cramér}, w}(n) - \Lambda_{\text{HB}, Q}(n)) e(R_0(n))| \lesssim Q^{-c}$$

for any interval J of length \sqrt{N} .

From Corollary 4.4 and the q -periodicity of R_0 , we have

$$\mathbb{E}_{n \in J} \Lambda_{\text{Cramér}, w}(n) e(R_0(n)) = \mathbb{E}_{n \in (\mathbb{Z}/q\mathbb{Z})^\times} e(R_0(n)) + O(Q^{-c})$$

(in fact, the error term is significantly better than this). Using the multiplicativity of the Ramanujan sums $c_q(\cdot)$ and the fact that $c_p(n) = (p-1)\mathbb{1}_{n=0 \pmod{p}} - \mathbb{1}_{n \neq 0 \pmod{p}}$, we have

$$\sum_{\ell|q} \frac{\mu(\ell)}{\varphi(\ell)} c_\ell(n) = \prod_{p|q} \left(1 - \frac{c_p(n)}{p-1}\right) = \mathbb{1}_{(n,q)=1} \frac{q}{\varphi(q)}.$$

We thus have

$$\mathbb{E}_{n \in J} \Lambda_{\text{Cramér}, w}(n) e(R_0(n)) = \sum_{\ell|q} \frac{\mu(\ell)}{\varphi(\ell)} \mathbb{E}_{n \in [q]} e(R_0(n)) c_\ell(n) + O(Q^{-c}).$$

Note that for any natural numbers ℓ, a, q with $\ell \nmid q$, by the geometric sum formula, we have

$$\mathbb{E}_{n \in J} c_\ell(n) \mathbb{1}_{n \equiv a \pmod{q}} = \sum_{r \in (\mathbb{Z}/\ell\mathbb{Z})^\times} \mathbb{E}_{n \in J} e\left(\frac{rn}{\ell}\right) \mathbb{1}_{n \equiv a \pmod{q}} \ll \ell^2 / \sqrt{N}.$$

Therefore, from equation (4.1) and the q -periodicity of $e(R_0(n))$, we have

$$\mathbb{E}_{n \in J} \Lambda_{\text{HB}, Q}(n) e(R_0(n)) = \sum_{\substack{\ell|q \\ \ell < Q}} \frac{\mu(\ell)}{\varphi(\ell)} \mathbb{E}_{n \in [q]} e(R_0(n)) c_\ell(n) + O(Q^{-c})$$

(again, a better error term is available here). Thus, by the triangle inequality, it suffices to show that

$$\sum_{\substack{\ell|q \\ \ell \geq Q}} \frac{\mu^2(\ell)}{\varphi(\ell)} |\mathbb{E}_{n \in [q]} e(R_0(n)) c_\ell(n)| \lesssim Q^{-c}.$$

By the divisor bound, q has at most $Q^{o(1)}$ factors, so it will suffice to establish the bound

$$|\mathbb{E}_{n \in [q]} e(R_0(n)) c_\ell(n)| \lesssim \varphi(\ell) Q^{-c}$$

for each square-free $\ell \mid q$ with $\ell \geq Q$. By the triangle inequality, it suffices to show that

$$\sum_{r \in (\mathbb{Z}/\ell\mathbb{Z})^\times} |\mathbb{E}_{n \in \mathbb{Z}/q\mathbb{Z}} e(R_0(n) - rn/\ell)| \lesssim \varphi(\ell) Q^{-c}.$$

However, from the Plancherel identity (or Bessel inequality) and the fact that $\ell \leq q$, one has

$$\sum_{r \in (\mathbb{Z}/\ell\mathbb{Z})^\times} |\mathbb{E}_{n \in \mathbb{Z}/q\mathbb{Z}} e(R_0(n) - rn/\ell)|^2 \leq \frac{\ell}{q} \leq 1,$$

and the claim follows from Cauchy–Schwarz (noting from the hypothesis $\ell \geq Q$ that $\varphi(\ell) \gtrsim Q^{1/2}$, say, so that $\varphi(\ell)^{1/2} \lesssim \varphi(\ell) Q^{-1/4}$). \square

5. Verifying the properties of the approximant

Recall the definition of Λ_N from equation (4.3). In this section, we verify the properties in equations (3.1), (3.2), (3.3) and (3.4) for Λ_N , and prove Proposition 3.4 and Theorem 3.2 concerning it.

Verifying equations (3.1), (3.2) and (3.3). The bound in equation (3.3) follows from Corollary 4.4, while the bound in equation (3.2) follows from Lemma 4.1. The bound in equation (3.1) follows, for instance, from [18, Theorem 1.1(ii)] (and could also be extracted from the earlier arguments in [17]). (Strictly speaking, the results in [18] were stated only for $C_0 = 10$, but an inspection of the arguments reveal that they also apply for larger choices of C_0 .)

Verifying equation (3.4). We need the following weighted analogue of [13, Proposition 6.21].

LEMMA 5.1. (*L^p improving*) *Let $Q \in \mathbb{Z}[n]$ be of degree $d \geq 1$. If $2 - c_d < p \leq 2$ for some sufficiently small $c_d > 0$, then*

$$\|\mathbb{E}_{n \in [N]} (\Lambda(n) + \Lambda_N(n)) f(\cdot + Q(n))\|_{\ell^2(\mathbb{Z})} \lesssim_Q N^{d/2-d/p} \|f\|_{\ell^p(\mathbb{Z})}$$

and also for the dual exponent $p' = p/(p-1)$, we have

$$\|\mathbb{E}_{n \in [N]} (\Lambda(n) + \Lambda_N(n)) f(\cdot + Q(n))\|_{\ell^{p'}(\mathbb{Z})} \lesssim_Q N^{d/p'-d/p} \|f\|_{\ell^p(\mathbb{Z})}. \quad (5.1)$$

The value of c_d here could be explicitly computed, but we do not attempt to optimize it here. After Lemma 5.1 has been proven, equation (5.1) together with the non-negativity of Λ_N immediately implies the required estimate in equation (3.4).

Proof. By interpolation (adjusting c_d as necessary), it suffices to show the second estimate in equation (5.1).

For any polynomial $Q(n) \in \mathbb{Z}[n]$, we define the averaging operators $A_N^{Q,0}$, $A_N^Q: \ell^p(\mathbb{Z}) \rightarrow \ell^p(\mathbb{Z})$ by the formulae

$$A_N^{Q,0} f(x) := \mathbb{E}_{n \in [N]} f(x + Q(n)) \Lambda(n), \quad A_N^Q f(x) := \mathbb{E}_{n \in [N]} f(x + Q(n)) \Lambda_N(n).$$

First, the operators $A_N^Q, A_N^{Q,0}$ are bounded on every $\ell^p(\mathbb{Z})$ thanks to equation (3.3) and the triangle inequality. With this notation, it suffices to show that

$$\begin{aligned}\|A_N^Q f\|_{\ell^{p'}(\mathbb{Z})} &\lesssim_Q N^{d/p'-d/p} \|f\|_{\ell^p(\mathbb{Z})}, \\ \|A_N^{Q,0} f\|_{\ell^{p'}(\mathbb{Z})} &\lesssim_Q N^{d/p'-d/p} \|f\|_{\ell^p(\mathbb{Z})}.\end{aligned}\quad (5.2)$$

We can write $A_N^Q = A_{N, \exp(\text{Log}^{1/C_0} N)}^Q$, where

$$A_{N,w}^Q f(x) := \mathbb{E}_{n \in [N]} f(x + Q(n)) \Lambda_{\text{Cramér}, w}(n).$$

On the one hand, from Lemma 4.1 and the results in [8] (see also [13, Proposition 6.21]), we have

$$\|A_{N,w}^Q f\|_{\ell^{p'}(\mathbb{Z})} \lesssim_Q N^{d/p'-d/p} \langle \text{Log } w \rangle \|f\|_{\ell^p(\mathbb{Z})} \quad (5.3)$$

for any $2 - c < p \leq 2$ (where $c > 0$ depends on d and can vary from line to line). On the other hand, from Lemma 4.5, we have

$$\mathbb{E}_{n \in [N]} (\Lambda_{\text{Cramér}, w} - \Lambda_{\text{Cramér}, z})(n) e(Q(n)) \lesssim_d z^{-c} \quad (5.4)$$

for any $1 \leq z \leq w \leq \exp(\text{Log}^{1/C_0} N)$.

By the Plancherel theorem, this implies that

$$\begin{aligned}\|A_{N,w}^Q f - A_{N,z}^Q f\|_{\ell^2(\mathbb{Z})} &= \left(\int_0^1 \left| \sum_{x \in \mathbb{Z}} (A_{N,w}^Q f - A_{N,z}^Q f)(x) e(\theta x) \right|^2 d\theta \right)^{1/2} \\ &\lesssim_d z^{-c} \left(\int_0^1 \left| \sum_{x \in \mathbb{Z}} f(x) e(\theta x) \right|^2 d\theta \right)^{1/2} \\ &\lesssim_d z^{-c} \|f\|_{\ell^2(\mathbb{Z})}.\end{aligned}$$

Interpolating (and reducing c as necessary), we see that if $2 - c \leq p \leq 2$, then

$$\|A_{N,w}^Q f - A_{N,z}^Q f\|_{\ell^{p'}(\mathbb{Z})} \lesssim_Q N^{d/p'-d/p} z^{-c} \|f\|_{\ell^p(\mathbb{Z})}$$

if $1 \leq z \leq w \leq \exp(\text{Log}^{1/C_0} N)$ is such that $w^{1/2} \leq z$. Summing this bound telescopically for suitable values of z, w , we conclude from the triangle inequality that

$$\|A_N^Q f - A_{N,1}^Q f\|_{\ell^{p'}(\mathbb{Z})} \lesssim_Q N^{d/p'-d/p} \|f\|_{\ell^p(\mathbb{Z})}.$$

Combining this with the $w = 1$ case of equation (5.3), we obtain the first estimate in equation (5.2).

The second estimate in equation (5.2) follows similarly, except that in the proof, we replace equation (5.3) with

$$\|A_N^{Q,0} f\|_{\ell^{p'}(\mathbb{Z})} \lesssim_Q N^{d/p'-d/p} \langle \text{Log } N \rangle \|f\|_{\ell^p(\mathbb{Z})}$$

and replace equation (5.4) with

$$\mathbb{E}_{n \in [N]} (\Lambda - \Lambda_{\text{Cramér}, z})(n) e(Q(n)) \lesssim_d z^{-c}$$

and use the first estimate in equation (5.2). □

Proof of Proposition 3.4. Arguing as in the proof of [13, Proposition 7.13], Proposition 3.4 reduces to establishing the symbol estimates

$$\left| \frac{\partial^{j_1}}{\partial \xi_1^{j_1}} \frac{\partial^{j_2}}{\partial \xi_2^{j_2}} M_0((\alpha_1, \xi_1), (\alpha_2, \xi_2)) \right| \lesssim_{C_3} 2^{O(\max(2^{\rho l}, s))} N^{j_1 + dj_2} \exp(-\text{Log}^c N)$$

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 $\xi_1 = O(2^s/N)$, $\xi_2 = O(2^{ds}/N^d)$, where the symbol M_0 is defined by the formula

$$M_0((\alpha_1, \xi_1), (\alpha_2, \xi_2)) := \mathbb{E}_{n \in [N]} e(\alpha_1 n + \alpha_2 P(n)) e(\xi_1 n + \xi_2 P(n)) \Lambda_N(n) \mathbb{1}_{n > N/2} - m_{\hat{\mathbb{Z}}^\times}(\alpha_1, \alpha_2) \tilde{m}_{N, \mathbb{R}}(\xi_1, \xi_2).$$

As in the proof of [13, Proposition 7.13], the function $n \mapsto e(\alpha_1 n + \alpha_2 P(n))$ is periodic of some period

$$q = O_\rho(2^{O(2^{\rho l})}). \quad (5.5)$$

In particular, from equation (3.19), one has

$$q \leq \exp(\text{Log}^{c_0} N)$$

and hence q divides W . So the function $\Lambda_N(n)$ vanishes outside of the primitive residue classes modulo q . Meanwhile, we have

$$m_{\hat{\mathbb{Z}}^\times}(\alpha_1, \alpha_2) = \mathbb{E}_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} e(\alpha_1 a + \alpha_2 P(a)).$$

By the triangle inequality, it thus suffices to show for each $a \in (\mathbb{Z}/q\mathbb{Z})^\times$ that

$$\left| \frac{\partial^{j_1}}{\partial \xi_1^{j_1}} \frac{\partial^{j_2}}{\partial \xi_2^{j_2}} \left(\mathbb{E}_{n \in [N]} e(\xi_1 n + \xi_2 P(n)) \Lambda_N(n) \mathbb{1}_{n=a \pmod{q}} \mathbb{1}_{n > N/2} - \frac{1}{\varphi(q)} \tilde{m}_{N, \mathbb{R}}(\xi_1, \xi_2) \right) \right| \lesssim_{C_3} 2^{O(\max(2^{\rho l}, s))} N^{j_1 + dj_2} \exp(-\text{Log}^c N).$$

Evaluating the derivatives, it suffices to show that

$$\left| \sum_{n \in [N] \setminus [N/2]} w(n) \mathbb{1}_{n=a \pmod{q}} \Lambda_N(n) - \frac{1}{\varphi(q)} \int_{N/2}^N w(t) dt \right| \lesssim_{C_3} 2^{O(\max(2^{\rho l}, s))} N^{j_1 + 2j_2 + 1} \exp(-\text{Log}^c N),$$

where

$$w(t) := e(\xi_1 t + \xi_2 P(t)) t^{j_1} P(t)^{j_2}.$$

The function w is smooth with a total variation of $O(2^{O(\max(2^{\rho l}, s))} N^{j_1 + 2j_2})$. Summing (or integrating) by parts as in [18, Lemma 2.2(iii)], it suffices to show that

$$\left| \sum_{n \in I} \left(\mathbb{1}_{n=a \pmod{q}} \Lambda_N(n) - \frac{1}{\varphi(q)} |I| \right) \right| \lesssim_{C_3} N \exp(-\text{Log}^c N)$$

for all intervals I in $[N, 2N]$. However, this follows from Corollary 4.4. □

Proof of Theorem 3.2. The last remaining task is to establish the single-scale estimate in Theorem 3.2. We first recall an application of the Peluse–Prendiville theory. □

PROPOSITION 5.2. (Unweighted 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 $f, g, h: \mathbb{Z} \rightarrow \mathbb{C}$ be supported on $[-N_0, N_0]$ with*

$$\|f\|_{\ell^\infty(\mathbb{Z})}, \|g\|_{\ell^\infty(\mathbb{Z})}, \|h\|_{\ell^\infty(\mathbb{Z})} \leq 1, \quad (5.6)$$

obeying the lower bound

$$|\langle \tilde{A}_{N,1}(f, g), h \rangle| \geq \delta N^d. \quad (5.7)$$

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

$$\|F\|_{\ell^\infty(\mathbb{Z})} \lesssim 1, \quad \|F\|_{\ell^1(\mathbb{Z})} \lesssim N^d \quad (5.8)$$

and with $\mathcal{F}_{\mathbb{Z}} F$ supported in the $O(\delta^{-O(1)}/N)$ -neighbourhood of some rational $a/b \bmod 1 \in \mathbb{Q}/\mathbb{Z}$ with $b = O(\delta^{-O(1)})$ such that

$$|\langle f, F \rangle| \gtrsim \delta^{O(1)} N^d. \quad (5.9)$$

Here, we use the inner product $\langle f, F \rangle := \sum_{n \in \mathbb{Z}} f(n) \overline{F(n)}$.

Proof. See [13, Proposition 6.6]. □

We now transfer this to the weighted setting, under an additional (mild) largeness hypothesis on δ .

PROPOSITION 5.3. (Weighted inverse theorem) *Let $N \geq 1$ and $\exp(-\text{Log}^{1/C_0} N) \leq \delta \leq 1$, and let N_0 be a quantity with $N_0 \sim N^d$. Let $f, g, h: \mathbb{Z} \rightarrow \mathbb{C}$ be supported on $[-N_0, N_0]$, obeying equation (5.6) and the lower bound*

$$|\langle \tilde{A}_{N, \Lambda_N}(f, g), h \rangle| \geq \delta N^d. \quad (5.10)$$

Then, the conclusions of Proposition 5.2 hold.

Proof. We may assume that N is sufficiently large depending on the fixed polynomial P , as the claim is easy to establish otherwise.

For any $1 \leq z \leq w \leq \exp(\text{Log}^{1/C_0} N)$, we have from Lemmas 4.5, 4.1 and [26, Theorem 4.1] (that is, equation (1.7)) that

$$|\langle \tilde{A}_{N, \Lambda_{\text{Cramér}, w} - \Lambda_{\text{Cramér}, z}}(f, g), h \rangle| \lesssim z^{-c} \langle \text{Log } w \rangle N^d.$$

In particular, we have

$$|\langle \tilde{A}_{N, \Lambda_{\text{Cramér}, w} - \Lambda_{\text{Cramér}, z}}(f, g), h \rangle| \lesssim z^{-c} N^d \quad (5.11)$$

for $z \in [w/2, w]$; summing dyadically using the triangle inequality, we conclude that

$$|\langle \tilde{A}_{N, \Lambda_N - \Lambda_{\text{Cramér}, w}}(f, g), h \rangle| \lesssim w^{-c} N^d$$

for any $1 \leq w \leq \exp(\text{Log}^{1/C_0} N)$.

The weight $\Lambda_{\text{Cramér}, w}$ is not quite of Type I form, so we now aim to swap it with the Heath-Brown weight $\Lambda_{\text{HB}, w}$. From Lemma 4.7, we have

$$\|\Lambda_{\text{Cramér}, w} - \Lambda_{\text{HB}, w}\|_{u^{d+1}[N]} \lesssim w^{-c}. \quad (5.12)$$

We would like to apply [26, Theorem 4.1] again, but we have the technical issue that $\Lambda_{\text{HB},w}$ does not quite have a good uniform bound, but is instead only controlled in the ℓ^k norm for arbitrarily large but finite k . However, from Lemma 4.6 (applied with sufficiently large k) and Chebyshev's inequality, for any small $\kappa > 0$ and $\varepsilon > 0$, we can find an approximation $\Lambda'_{\text{HB},w}$ to $\Lambda_{\text{HB},w}$ with

$$\|\Lambda_{\text{HB},w} - \Lambda'_{\text{HB},w}\|_{\ell^1[N]} \leq \kappa \quad \text{and} \quad \Lambda'_{\text{HB},w}(n) = O_\varepsilon(\kappa^{-\varepsilon} \langle \text{Log } w \rangle^{O_\varepsilon(1)}). \quad (5.13)$$

We can use the ℓ^1 norm to control the u^{d+1} norm; hence, by equation (5.12) and the triangle inequality,

$$\|\Lambda_{\text{Cramér},w} - \Lambda'_{\text{HB},w}\|_{u^{d+1}[N]} \lesssim \kappa + w^{-c}. \quad (5.14)$$

Now, we can apply [26, Theorem 4.1] (and Lemma 4.1) to conclude that

$$|\langle \tilde{A}_{N,\Lambda_{\text{Cramér},w} - \Lambda'_{\text{HB},w}}(f, g), h \rangle| \lesssim_\varepsilon \langle \text{Log } w \rangle^{O_\varepsilon(1)} (\kappa^c + \kappa^{-\varepsilon} w^{-c}) N^d.$$

Finally, from the triangle inequality and Cauchy–Schwarz, we can crudely bound

$$|\langle \tilde{A}_{N,\Lambda'_{\text{HB},w} - \Lambda_{\text{HB},w}}(f, g), h \rangle| \lesssim \kappa N^d.$$

Putting this all together, choosing ε to be sufficiently small and κ to be a small multiple of w^{-c} for a suitable c , we conclude that

$$|\langle \tilde{A}_{N,\Lambda_N - \Lambda_{\text{HB},w}}(f, g), h \rangle| \lesssim w^{-c} N^d$$

for any $1 \leq w \leq \exp(\text{Log}^{1/C_0} N)$. In particular, from equation (5.10), we now have

$$|\langle \tilde{A}_{N,\Lambda_{\text{HB},w}}(f, g), h \rangle| \gtrsim \delta N^d$$

for some $1 \leq w \lesssim \delta^{-O(1)}$. Expanding equation (4.1) and using the triangle inequality and crude bounds, we conclude that

$$|\langle \tilde{A}_{N,e(-r \cdot /q)}(f, g), h \rangle| \gtrsim \delta^{O(1)} N^d$$

for some $1 \leq r \leq q \lesssim \delta^{-O(1)}$. However, observe the identity

$$\langle \tilde{A}_{N,e(-r \cdot /q)}(f, g), h \rangle = \langle \tilde{A}_{N,1}(e(-r \cdot /q)f, g), e(-r \cdot /q)h \rangle.$$

We can thus apply Proposition 5.2 to conclude that

$$|\langle e(-r \cdot /q)f, F \rangle| \gtrsim \delta^{O(1)} N^d$$

for some function F obeying the conclusions of that proposition. Transferring the plane wave $e(-r \cdot /q)$ from f to F , we obtain the claim (noting that the denominator b will remain acceptably under control since $q \lesssim \delta^{-O(1)}$). \square

If we now repeat the arguments of [13, §6.1], using Proposition 5.3 and Lemma 5.1 in place of [13, Proposition 6.6] and [13, Proposition 6.21], respectively, inserting the weights Λ_N in the averaging operators in the obvious fashion, we obtain case (i) of Theorem 3.2. To handle case (ii), we need the following variant of Proposition 5.3.

PROPOSITION 5.4. (Weighted inverse theorem for g) *Under the hypotheses of Proposition 5.3, there exists a function $G \in \ell^2(\mathbb{Z})$ with*

$$\|G\|_{\ell^\infty(\mathbb{Z})} \lesssim 1, \quad \|G\|_{\ell^1(\mathbb{Z})} \lesssim N^d \quad (5.15)$$

and with $\mathcal{F}_{\mathbb{Z}}G$ supported in the $O(\delta^{-O(1)}/N^d)$ -neighbourhood of some rational $a/b \bmod 1 \in \mathbb{Q}/\mathbb{Z}$ with $b = O(\delta^{-O(1)})$ such that

$$|\langle g, G \rangle| \gtrsim \delta^{O(1)} N^d. \quad (5.16)$$

However, this can be derived from [13, Proposition 6.26] in precisely the same way Proposition 5.3 was derived from [13, Proposition 6.6]. By repeating the remaining arguments of [13, §6.2], one obtains case (ii) of Theorem 3.2.

6. Remarks

6.1. *Peluse's inverse theorem for the primes.* As is clear from the previous sections, Peluse's inverse theorem [23] was an important ingredient in the proof of the unweighted bilinear ergodic theorem in [13]. In the course of proving Theorem 1.3, we essentially needed a version of this inverse theorem where one of the variables was weighted by the approximant Λ_N ; see Proposition 5.3. It is natural to ask if one can also obtain a version of Peluse's inverse theorem with the von Mangoldt weight Λ . We record here how such a result quickly follows from the arguments used to prove Proposition 5.3.

THEOREM 6.1. (Peluse's inverse theorem with prime weight) *Let $k, d \in \mathbb{N}$ and $A > 0$. Let $N \geq 2$, $(\log N)^{-A} \leq \delta \leq 1$ and $N_0 \sim N^d$. Let P_1, \dots, P_k be polynomials with integer coefficients of distinct degrees, with maximal degree d . Let $h, f_1, \dots, f_k: \mathbb{Z} \rightarrow \mathbb{C}$ be functions bounded in modulus by 1 and supported on $[-N_0, N_0]$. Suppose that*

$$\left| \sum_{x \in \mathbb{Z}} \mathbb{E}_{n \in [N]} \Lambda(n) h(x) f_1(x + P_1(n)) \cdots f_k(x + P_k(n)) \right| \geq \delta N^d. \quad (6.1)$$

Then, either $N_0 \lesssim_{P_1, \dots, P_k} \delta^{-O_d(1)}$ or there exists a positive integer $q \lesssim_{P_1, \dots, P_k} \delta^{-O_d(1)}$ and $\delta^{O_d(1)} N \lesssim_{P_1, \dots, P_k} N' \leq N$ such that

$$\frac{1}{N^d} \left| \sum_{x \in \mathbb{Z}} \mathbb{E}_{m \in [N']} f_1(x + qm) \right| \gtrsim_{A, P_1, \dots, P_k} \delta^{O_d(1)}.$$

Proof. Fix P_1, \dots, P_k ; we allow all implied constants to depend on them. Define the polynomial averaging operator

$$T_{N, \theta}(h, f_1, \dots, f_k) := \sum_{x \in \mathbb{Z}} \mathbb{E}_{n \in [N]} \theta(n) h(x) f_1(x + P_1(n)) \cdots f_k(x + P_k(n)).$$

Let $w_0 = \delta^{-C_d}$ for a large enough constant C_d . We claim that

$$T_{N, \Lambda - \Lambda_N}(h, f_1, \dots, f_k) \lesssim_A (\log N)^{-A}, \quad (6.2)$$

and

$$T_{N, \Lambda_N - \Lambda_{\text{Cramér}, w_0}}(h, f_1, \dots, f_k) \lesssim \delta^2 \quad (6.3)$$

and

$$T_{N, \Lambda_{\text{Cramér}, w_0} - \Lambda_{\text{HB}, w_0}}(h, f_1, \dots, f_k) \lesssim \delta^2. \quad (6.4)$$

After we have these three estimates, we conclude from equation (6.1) and linearity that

$$|T_{N, \Lambda_{\text{HB}, w_0}}(h, f_1, \dots, f_k)| \gtrsim \delta.$$

By equations (4.1) and (4.2), the function Λ_{HB, w_0} is a linear combination, with 1-bounded coefficients, of $O(w_0^3)$ indicators of arithmetic progressions of common difference at most w_0 . Hence, crudely using the triangle inequality, we obtain

$$|T_{N, \mathbb{I}_{a(q')}}(h, f_1, \dots, f_k)| \gtrsim \delta^{O_d(1)}$$

for some $1 \leq a \leq q' \lesssim \delta^{-O_d(1)}$. However, now the claim of the theorem follows from [23, Theorem 3.3] after making a change of variables.

We are left with showing equations (6.2), (6.3) and (6.4). The estimate in equation (6.2) follows immediately from [26, Theorem 4.1] and equation (3.1). The estimate in equation (6.3) follows by using Lemmas 4.5, 4.1 and [26, Theorem 4.1] to obtain

$$T_{N, \Lambda_{\text{Cramér}, w} - \Lambda_{\text{Cramér}, z}}(h, f_1, \dots, f_k) \lesssim w^{-c_d}$$

for some $c_d > 0$ and any $z \in [w/2, w]$, $1 \leq w \leq \exp((\log N)^{1/10})$, and then summing this dyadically. For proving equation (6.4), note that from equation (5.14) and [26, Theorem 4.1], we have for any $\kappa > 0$, $\varepsilon > 0$, the bound

$$T_{N, \Lambda_{\text{Cramér}, w_0} - \Lambda'_{\text{HB}, w_0}}(h, f_1, \dots, f_k) \lesssim_{\varepsilon} (\text{Log } w_0)^{O_{\varepsilon}(1)} (\kappa^{c'_d} + \kappa^{-\varepsilon} w_0^{-c'_d}) N^d,$$

with $\Lambda'_{\text{HB}, w_0}$ obeying equation (5.13). However, from equation (5.13) and the triangle inequality, we now obtain equation (6.4) by taking $\varepsilon > 0$ small enough and $\kappa = w_0^{-c}$ for a small enough constant c (depending on d). This was enough to complete the proof. \square

6.2. Siegel zeroes. In this subsection, we mention an alternative approach to Theorem 1.3 based on working with Siegel zeroes. This approach is somewhat more complicated than that implemented above and we shall only sketch it very briefly, leaving the details to the interested reader.

The place in the proof of Theorem 1.3 where passing from the von Mangoldt function Λ to the approximant Λ_N avoided dealing with Siegel zeroes is Proposition 3.4, so we begin by sketching how a variant of Proposition 3.4 can be proven for the weight Λ .

We say that a modulus $q \geq 2$ is exceptional if there exists a non-principal real Dirichlet character $\chi_q \pmod{q}$ such that $L(s, \chi_q)$ has a real zero $\beta_q > 1 - c_0/(\log q)$, where c_0 is some small absolute constant. We call the corresponding character χ_q an exceptional character and we call β_q a Siegel zero. For any given exceptional q , the character χ_q and Siegel zero β_q are uniquely determined.

For exceptional characters χ_q , we define the arithmetic symbol

$$m_{\hat{\mathbb{Z}}^\times, \chi_q} \left(\frac{a_1}{q} \pmod{1}, \frac{a_2}{q} \pmod{1} \right) := \frac{1}{\phi(q)} \sum_{r \in (\mathbb{Z}/q\mathbb{Z})^\times} e \left(\frac{a_1 r}{q} + \frac{a_2 P(r)}{q} \right) \chi_q(r)$$

and the (weighted) continuous multiplier

$$\tilde{m}_{N,\mathbb{R},\chi_q}(\xi_1, \xi_2) := \int_{1/2}^1 e(\xi_1 t + \xi_2 P(t)) t^{\beta_q-1} dt,$$

where $\beta_q \in (0, 1)$ is the Siegel zero. Then, if we replace in equation (3.18),

$$\begin{aligned} & B_{(\eta \leq -\log N+s \otimes \eta \leq -d \log N+ds)}^{l_1, l_2, m_{\hat{\mathbb{Z}}^\times}} \tilde{m}_{N,\mathbb{R}} \\ & \longrightarrow B_{(\eta \leq -\log N+s \otimes \eta \leq -d \log N+ds)}^{l_1, l_2, m_{\hat{\mathbb{Z}}^\times}} \tilde{m}_{N,\mathbb{R}} + \sum_{q \text{ exceptional}} B_{(\eta \leq -\log N+s \otimes \eta \leq -d \log N+ds)}^{l_1, l_2, m_{\hat{\mathbb{Z}}^\times}, \chi_q} \tilde{m}_{N,\mathbb{R},\chi_q}, \end{aligned}$$

the conclusion of Proposition 3.4 holds with the von Mangoldt weight Λ in place of Λ_N . This follows from essentially the same proof as in §5, but using the Landau–Page theorem [20, Corollary 11.10] in place of Corollary 4.4.

In the large-scale regime, the error bounds arising from the Siegel–Walfisz theorem remove the need for the above approximation; in the small-scale regime,

$$\{N \in \mathbb{D}: 2^{u^{O(1/(C_0 \rho))}}} \leq N \leq 3^{C_0 \cdot 2^u}\},$$

further analysis is required to reduce matters to the two-parameter Rademacher–Menshov inequality.

The first observation is the classical fact that there is at most one exceptional character at each dyadic scale:

$$|\{q \in (2^j, 2^{j+1}]: q \text{ exceptional}\}| \leq 1. \quad (6.5)$$

We let q_j denote the unique exceptional modulus in $(2^j, 2^{j+1}]$ and abbreviate $\beta_j = \beta_{q_j}$.

We then introduce a dyadic decomposition

$$\sum_{q \text{ exceptional}} B_{(\eta \leq -\log N+s \otimes \eta \leq -d \log N+ds)}^{l_1, l_2, m_{\hat{\mathbb{Z}}^\times}, \chi_q} \tilde{m}_{N,\mathbb{R},\chi_q} = \sum_{j \leq 2^{\rho l}} C_{N,j}(f, g),$$

where

$$\begin{aligned} & C_{N,j}(f, g)(x) \\ &= \int_{1/2}^1 \left(\int_{\mathbb{T}^2} \sum_{(a_1/q_j, a_2/q_j): h(a_i/q_j)=2^{l_i}} m_{\hat{\mathbb{Z}}^\times, \chi_{q_j}}(a_1/q_j, a_2/q_j) e(a_1 x/q_j + a_2 x/q_j) \right. \\ & \quad \times (\hat{f}(\xi_1 + a_1/q_j) \cdot \varphi(2^u \xi_1) \cdot e(\xi_1 N t)) \\ & \quad \times (\hat{g}(\xi_2 + a_2/q_j) \cdot \varphi(2^{du} \xi_2) \cdot e(\xi_2 P(N t))) e(\xi_1 x + \xi_2 x) \cdot N^{\beta_j-1} t^{\beta_j-1} d\xi_1 d\xi_2 \Big) dt. \end{aligned}$$

The key novelty then derives from proving the following modified Rademacher–Menshov-type inequality, similar to [13, Lemma 8.2].

LEMMA 6.2. *Let V, W be normed vector spaces, K, J be two positive integers and let $0 < q < \infty$. Let $B_j: V \times W \rightarrow L^q(X)$ be a family of bilinear operators for $j \in [J]$. Let $\{f_k^j\}, \{g_k^j\}$ be sets of functions with $f_k^j \in V$ and $g_k^j \in W$ for $k \in [K]$ and $j \in [J]$. Then,*

$$\left\| V^2 \left(\sum_{j \in [J]} B_j(f_k^j, g_k^j) : k \in [K] \right) \right\|_{L^q(X)} \\ \lesssim_q \langle \log K \rangle^{O_q(1)} \sup_{\epsilon_k^j, \epsilon_k^j \in \{\pm 1\}} \left\| \sum_{j \in [J]} B_j \left(\sum_{k \in [K]} \epsilon_k^j (f_k^j - f_{k-1}^j), \sum_{k \in [K]} \epsilon_k^j (g_k^j - g_{k-1}^j) \right) \right\|_{L^q(X)}.$$

This result may be of independent interest, so we provide a brief proof.

Proof. Set $a_{k_1, k_2} = \sum_{j \in [J]} B_j(f_{k_1}^j, g_{k_2}^j)$. By [13, Lemma 8.1], we have

$$V^2 \left(\sum_{j \in [J]} B_j(f_k^j, g_k^j) : k \in [K] \right) \lesssim \sum_{\substack{M_1, M_2 < K \\ \mathcal{M}_{1, M_2}: \text{dyadic}}} \left\| \Delta \sum_{j \leq J} B_j(f_{M_1 n_1}^j, g_{M_2 n_2}^j) \right\|_{\ell^2(n_1, n_2)},$$

where

$$\Delta \sum_{j \in [J]} B_j(f_{M_1 n_1}^j, g_{M_2 n_2}^j) = \sum_{j \in [J]} B_j(f_{M_1 n_1}^j, g_{M_2 n_2}^j) - \sum_{j \in [J]} B_j(f_{(n_1-1)M_1}^j, g_{M_2 n_2}^j) \\ - \sum_{j \in [J]} B_j(f_{M_1 n_1}^j, g_{(n_2-1)M_2}^j) + \sum_{j \in [J]} B_j(f_{(n_1-1)M_1}^j, g_{(n_2-1)M_2}^j).$$

Taking

$$\tilde{f}_{M_1 n_1} = f_{M_1 n_1} - f_{(n_1-1)M_1}, \quad \tilde{g}_{M_2 n_2} = g_{M_2 n_2} - g_{(n_2-1)M_2},$$

we need to bound

$$\langle \log K \rangle^{O_q(1)} \sup_{\substack{M_1, M_2 < K \\ \text{dyadic}}} \left\| \left(\sum_{\substack{n_1 < k/M_1 \\ n_2 < k/M_2}} \left| \sum_{j \in [J]} B_j(\tilde{f}_{M_1 n_1}^j, \tilde{g}_{M_2 n_2}^j) \right|^2 \right)^{1/2} \right\|_{L^q(X)}. \quad (6.6)$$

Applying Khintchine's inequality

$$\left(\sum_n |a_n|^2 \right)^{1/2} = \left(\mathbb{E}_{\epsilon_n \in \pm 1} \left| \sum_n \epsilon_n a_n \right|^2 \right)^{1/2} \sim_q \left(\mathbb{E}_{\epsilon_n \in \pm 1} \left| \sum_n \epsilon_n a_n \right|^q \right)^{1/q},$$

we arrive at the following chain of inequalities:

$$\left\| V^2 \left(\sum_{j \in [J]} B_j(f_k^j, g_k^j) : k \in [K] \right) \right\|_{L^q(X)} \\ \lesssim \langle \log K \rangle^{O_q(1)} \sup_{M_1, M_2} \left\| \left(\mathbb{E}_{\epsilon_{n_2} \in \pm 1} \sum_{n_1} \left| \sum_{n_2} \sum_{j \in [J]} \epsilon_{n_2} B_s(\tilde{f}_{M_1 n_1}^j, \tilde{g}_{M_2 n_2}^j) \right|^2 \right)^{1/2} \right\|_{L^q(X)} \\ \lesssim \langle \log K \rangle^{O_q(1)} \sup_{M_1, M_2} \left\| \left(\mathbb{E}_{\epsilon_{n_1}, \epsilon_{n_2} \in \pm 1} \left| \sum_{n_1} \sum_{n_2} \sum_{j \in [J]} \epsilon_{n_1} \epsilon_{n_2} B_j(\tilde{f}_{M_1 n_1}^j, \tilde{g}_{M_2 n_2}^j) \right|^2 \right)^{1/2} \right\|_{L^q(X)} \\ \lesssim_q \langle \log K \rangle^{O_q(1)} \sup_{M_1, M_2} \left\| \left(\mathbb{E}_{\epsilon_{n_1}, \epsilon_{n_2} \in \pm 1} \left| \sum_{n_1} \sum_{n_2} \sum_{j \in [J]} \epsilon_{n_1} \epsilon_{n_2} B_j(\tilde{f}_{M_1 n_1}^j, \tilde{g}_{M_2 n_2}^j) \right|^q \right)^{1/q} \right\|_{L^q(X)} \\ \lesssim_q \langle \log K \rangle^{O_q(1)} \sup_{M_1, M_2, \epsilon_{n_1}, \epsilon_{n_2}} \left\| \sum_{n_1} \sum_{n_2} \sum_{j \in [J]} \epsilon_{n_1} \epsilon_{n_2} B_j(\tilde{f}_{M_1 n_1}^j, \tilde{g}_{M_2 n_2}^j) \right\|_{L^q(X)}.$$

By bilinearity, we may consolidate

$$\sum_{n_1} \sum_{n_2} \sum_{j \in [J]} \epsilon_{n_1} \epsilon_{n_2} B_j(\tilde{f}_{M_1 n_1}^j, \tilde{g}_{M_2 n_2}^j) = \sum_{j \in [J]} B_j \left(\sum_{n_1} \epsilon_{n_1} \tilde{f}_{M_1 n_1}^j, \sum_{n_2} \epsilon_{n_2} \tilde{g}_{M_2 n_2}^j \right);$$

putting everything together,

$$\begin{aligned} & \left\| V^2 \left(\sum_{j \in [J]} B_j(f_k^j, g_k^j) : k \in [K] \right) \right\|_{L^q(X)} \\ & \lesssim (\text{Log } K)^{O_q(1)} \sup_{\substack{M_1, M_2 \\ \epsilon_{n_1}, \epsilon_{n_2}}} \left\| \sum_{j \in [J]} B_j \left(\sum_{n_1} \epsilon_{n_1} (f_{M_1 n_1}^j - f_{(n_1-1)M_1}^j), \right. \right. \\ & \quad \left. \left. \sum_{n_2} \epsilon_{n_2} (g_{M_2 n_2}^j - g_{(n_2-1)M_2}^j) \right) \right\|_{L^q(X)}, \end{aligned}$$

and so we get the result upon telescoping e.g.

$$\begin{aligned} \epsilon_{n_1} (f_{M_1 n_1}^j - f_{(n_1-1)M_1}^j) &= \sum_{(n_1-1)M_1 < k \leq M_1 n_1} \epsilon_{n_1} (f_k^j - f_{k_1}^j) \\ &=: \sum_{(n_1-1)M_1 < k \leq M_1 n_1} \epsilon_k^j (f_k^j - f_{k-1}^j). \end{aligned} \quad \square$$

6.3. Breaking duality. We briefly remark that one may establish Theorem 1.3 with r -variation restricted to the range $r > 2 + \epsilon$ for exponents $p_1, p_2 > 1$ that satisfy

$$1 < \frac{1}{p} := \frac{1}{p_1} + \frac{1}{p_2} < 1 + \epsilon',$$

where $\epsilon' > 0$ is sufficiently small in terms of ϵ ; hence, going beyond the duality range.

The single-scale estimate

$$\|A_{N;\Lambda;X}(f, g)\|_{L^p(X)} \lesssim \|f\|_{L^{p_1}(X)} \|g\|_{L^{p_2}(X)} \quad (6.7)$$

anchors the argument; equation (6.7) follows from Hölder's inequality and the improving estimate Lemma 5.1, as per [13, Lemma 11.1]. With equation (6.7) in hand, the proof of [13, Proposition 11.4] can be formally reproduced, with only notational changes arising. We leave the details to the interested reader.

6.4. Sharpness of the variational result. The unboundedness of the quadratic variation along polynomial orbits, namely [13, Proposition 12.1], extends to our context.

PROPOSITION 6.3. *Let $P \in \mathbb{Z}[n]$ be a non-constant polynomial and let $0 < p \leq \infty$. Let $I \subset \mathbb{N}$ be an infinite set. Then, for every $C > 0$, there exists a measure-preserving system (X, μ, T) of total measure 1 and a 1-bounded $f \in L^\infty(X)$ so that*

$$\|(\mathbb{E}_{p \in [N]} T^{P(p)} f)_{N \in I}\|_{L^p(X; V^2)} \geq C.$$

We shall leave the details of the proof of this proposition to the interested reader as it is similar to the proof of [13, Proposition 12.1]. The key additional observation is the equidistribution of

$$p \mapsto (\alpha_1 \cdot P(p), \dots, \alpha_K \cdot P(p)) \subset \mathbb{T}^K$$

over the primes whenever $\alpha_1, \dots, \alpha_K$ are \mathbb{Q} -linearly independent and $P \in \mathbb{Z}[n]$ is a non-constant polynomial (which follows from Weyl's criterion and a standard exponential sum estimate for polynomials of primes; see e.g. [17, Theorem 1.3]).

To see why this implies the sharpness of the range of the variational estimate in Theorem 1.3, one may employ the convexity arguments of [19, §5], taking into account [19, Proposition 4.1], to obtain the lower bound

$$\|(\mathbb{E}_{p \in [N]} T^{P(p)} f)_{N \in I}\|_{L^p(X; V^2)} \leq \|(\mathbb{E}_{n \in [N]} \Lambda(n) \cdot T^{P(n)} f)_{N \in I}\|_{L^p(X; V^2)} + O(1).$$

6.5. Continuous extensions. From the perspective of density, the primes are 'full/dimensional', with a very 'Fourier-uniform' measure, Λ . A natural question concerns establishing a continuous analogue of Theorem 1.3, namely the existence of a measure ν supported on $[0, 1]$, with (say) full Fourier dimension,

$$|\hat{\nu}(\xi)| \lesssim (1 + |\xi|)^{o(1)-1/2}$$

so that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \int_0^N f(x-t)g(x-P(t)) \, d\nu(t), \quad d = \deg(P) \geq 2$$

exists almost everywhere whenever $f \in L^{p_1}(\mathbb{R})$ and $g \in L^{p_2}(\mathbb{R})$ with $p_1, p_2 > 1$ and $1/p_1 + 1/p_2 \leq 1$. The key point is establishing a suitable Sobolev inequality, namely

$$\left\| \frac{1}{N} \int_0^N f(x-t)g(x-P(t)) \, d\nu(t) \right\|_{L^1([0, CN^d])} \lesssim (2^{-cl} + O_A((\log N)^{-A}))N^d$$

for some $c > 0$, whenever $|f|, |g| \leq 1$, and \hat{f} vanishes on $\{|\xi| \lesssim 2^l/N\}$ and/or \hat{g} vanishes on $\{|\xi| \lesssim 2^l/N^d\}$.

Estimates of this form in the unweighted setting go back to [1], with the strongest estimates recently established by one of us as part of a much more general phenomenon, see [12]. This approach relies on PET induction, which suggests that certain Gowers-uniformity conditions might need to be imposed on ν ; it is unclear how this might interact with dimension, so we leave the problem to the interested reader.

Acknowledgments. We thank the referee for careful reading of the paper. B.K. is supported by an EPSRC New Investigators grant and an ERC Starting grant. T.T. is supported by NSF grant DMS-2347850. J.T. is supported by European Union's Horizon Europe research and innovation programme under Marie Skłodowska-Curie grant agreement No. 101058904 and ERC grant agreement No. 101162746, and Academy of Finland grant No. 362303.

REFERENCES

- [1] J. Bourgain. A nonlinear version of Roth's theorem for sets of positive density in the real line. *J. Anal. Math.* **50** (1988), 169–181.
- [2] J. Bourgain. An approach to pointwise ergodic theorems. *Geometric Aspects of Functional Analysis (1986/87) (Lecture Notes in Mathematics, 317)*. Springer, Berlin, 1988, pp. 204–223.

- [3] N. Frantzikinakis. Some open problems on multiple ergodic averages. *Bull. Hellenic Math. Soc.* **60** (2016), 41–90.
- [4] N. Frantzikinakis, B. Host and B. Kra. Multiple recurrence and convergence for sequences related to the prime numbers. *J. Reine Angew. Math.* **611** (2007), 131–144.
- [5] B. Green and T. Tao. An inverse theorem for the Gowers $U^3(G)$ norm. *Proc. Edinb. Math. Soc.* (2) **51** (2008), 73–153.
- [6] B. Green and T. Tao. Linear equations in primes. *Ann. of Math.* (2) **171** (2010), 1753–1850.
- [7] B. Green and T. Tao. The Möbius function is strongly orthogonal to nilsequences. *Ann. of Math.* (2) **175** (2012), 541–566.
- [8] R. Han, V. Kovač, M. T. Lacey, J. Madrid and F. Yang. Improving estimates for discrete polynomial averages. *J. Fourier Anal. Appl.* **26** (2020), 42.
- [9] D. R. Heath-Brown. The Ternary Goldbach problem. *Rev. Mat. Iberoam.* **1** (1985), 45–59.
- [10] A. Ionescu and S. Wainger. L^p boundedness of discrete singular Radon transforms. *J. Amer. Math. Soc.* **19** (2006), 357–383.
- [11] H. Iwaniec and E. Kowalski. *Analytic Number Theory (AMS Colloquium Publications, 53)*. American Mathematical Society, Providence, RI, 2004.
- [12] B. Krause, M. Mirek, S. Peluse and J. Wright. Polynomial progressions in topological fields. *Forum Math. Sigma* **12** (2024), e106.
- [13] B. Krause, M. Mirek and T. Tao. Pointwise ergodic theorems for non-conventional bilinear polynomial averages. *Ann. of Math.* (2) **195** (2022), 997–1109.
- [14] J. Leng. Improved quadratic Gowers uniformity for the Möbius function. *Preprint*, 2023, [arXiv:2212.09635](https://arxiv.org/abs/2212.09635).
- [15] J. Leng. Efficient equidistribution of nilsequences. *Preprint*, 2024, [arXiv:2312.10772](https://arxiv.org/abs/2312.10772).
- [16] J. Leng, A. Sah and M. Sawhney. Quasipolynomial bounds on the inverse theorem for the Gowers $U^{s+1}[N]$ -norm. *Preprint*, 2024, [arXiv:2402.17994](https://arxiv.org/abs/2402.17994).
- [17] K. Matomäki and X. Shao. Discorrelation between primes in short intervals and polynomial phases. *Int. Math. Res. Not. IMRN* **2021** (2021), 12330–12355.
- [18] K. Matomäki, X. Shao, T. Tao and J. Teräväinen. Higher uniformity of arithmetic functions in short intervals I. All intervals. *Forum Math. Pi* **11** (2023), e29.
- [19] M. Mirek, B. Trojan and P. Zorin-Kranich. Variational estimates for averages and truncated singular integrals along the prime numbers. *Trans. Amer. Math. Soc.* **369**(8) (2017), 5403–5423.
- [20] H. L. Montgomery and R. C. Vaughan. *Multiplicative Number Theory. I. Classical Theory (Cambridge Studies in Advanced Mathematics, 97)*. Cambridge University Press, Cambridge, 2007.
- [21] R. Nair. On polynomials in primes and J. Bourgain’s circle method approach to ergodic theorems. *Ergod. Th. & Dynam. Sys.* **11**(3) (1991), 485–499.
- [22] R. Nair. On polynomials in primes and J. Bourgain’s circle method approach to ergodic theorems. II. *Studia Math.* **105**(3) (1993), 207–233.
- [23] S. Peluse. Bounds for sets with no polynomial progressions. *Forum Math. Pi* **8** (2020), e16.
- [24] S. Peluse and S. Prendiville. A polylogarithmic bound in the nonlinear Roth theorem. *Int. Math. Res. Not. IMRN* **2022** (2022), 5658–5684.
- [25] T. Tao and J. Teräväinen. Quantitative bounds for Gowers uniformity of the Möbius and von Mangoldt functions. *J. Eur. Math. Soc. (JEMS)* **27** (2025), 1321–1384.
- [26] J. Teräväinen. Pointwise convergence of ergodic averages with Möbius weight. *Preprint*, 2024, [arXiv:2401.03174](https://arxiv.org/abs/2401.03174).
- [27] M. Wierdl. Pointwise ergodic theorem along the prime numbers. *Israel J. Math.* **64**(3) (1988), 315–336.
- [28] T. D. Wooley and T. D. Ziegler. Multiple recurrence and convergence along the primes. *Amer. J. Math.* **134** (2012), 1705–1732.