Math. Res. Lett.
Volume 29, Number 6, 1767–1791, 2022

Endpoint ℓ^r improving estimates for prime averages

Michael T. Lacey, Hamed Mousavi, and Yaghoub Rahimi

Let Λ denote von Mangoldt's function, and consider the averages

$$A_N f(x) = \frac{1}{N} \sum_{1 \le n \le N} f(x-n)\Lambda(n).$$

We prove sharp ℓ^p-improving for these averages, and sparse bounds for the maximal function. The simplest inequality is that for sets $F, G \subset [0, N]$ there holds

$$N^{-1} \langle A_N \mathbf{1}_F, \mathbf{1}_G \rangle \ll \frac{|F| \cdot |G|}{N^2} \Bigl(\operatorname{Log} \frac{|F| \cdot |G|}{N^2} \Bigr)^t,$$

where $t = 2$, or assuming the Generalized Riemann Hypothesis, $t = 1$. The corresponding sparse bound is proved for the maximal function $\sup_N A_N \mathbf{1}_F$. The inequalities for $t = 1$ are sharp. The proof depends upon the Circle Method, and an interpolation argument of Bourgain.

1. Introduction

We consider discrete averages over the prime integers. The averages are weighted by the von Mangoldt function.

$$A_N f(x) = \frac{1}{N} \sum_{1 \le n \le N} f(x-n)\Lambda(n)$$

$$\Lambda(n) = \begin{cases} \log(p) & n = p^a, p \text{ prime} \\ 0 & \text{Otherwise.} \end{cases}$$

Our interest is in *scale free* ℓ^r improving estimates for these averages. The question presents itself in different forms.

For an interval I in the integers and function $f : I \to \mathbb{C}$, set

$$\langle f \rangle_{I,r} = \Big[|I|^{-1} \sum_{x \in I} |f(x)|^r \Big]^{1/r}.$$

If $r = 1$, we will suppress the index in the notation. And, set $\operatorname{Log} x = 1 + |\log x|$, for $x > 0$.

The kind of estimate we are interested in takes the the following form, in the simplest instance. What is the 'smallest' function $\psi : [0,1] \to [1, \infty)$ so that for all integers N and indicator functions $f, g : I \to \{0,1\}$, there holds

$$N^{-1} \langle A_N f, g \rangle \leq \langle f \rangle_I \langle g \rangle_I \psi(\langle f \rangle_I \langle g \rangle_I).$$

That is, the right hand side is independent of N, making it scale-free. We specified that f, g be indicator functions as that is sometimes the sharp form of the inequality. Of course it is interesting for arbitrary functions, but the bound above is not homogeneous, so not the most natural estimate in that case.

The points of interest in these two results arises from, on the one hand, the distinguished role of the prime integers. And, on the other, endpoint results are significant interest in Harmonic Analysis, as the techniques which apply are the sharpest possible. In this instance, the sharp methods depend very much on the prime numbers.

For the primes, we expect that the Riemann Hypothesis to be relevant. We state unconditional results, and those that depend upon the Generalized Riemann Hypothesis (GRH). Note that according to GRH all zeroes in the critical strip $0 < Re(s) < 1$ of an arbitrary $L-$function $L(f, s)$ are on the critical line $Re(s) = \frac{1}{2}$. Under GRH, the primes are equitably distributed mod q, with very good error bounds. Namely,

$$\psi(x, q, a) = \sum_{\substack{n < x \\ n \equiv a \pmod q}} \Lambda(n) = \frac{x}{\phi(q)} + O(x^{\frac{1}{2}} \log^2(q)). \tag{1.1}$$

Theorem 1.2. *There is a constant C so that this holds. For integers $N > 30$, and interval I of length N, the following inequality holds for all functions $f = \mathbf{1}_F$ and $g = \mathbf{1}_G$ with $F, G \subset I$*

$$N^{-1} \langle A_N f, g \rangle \leq C \langle f \rangle_I \langle g \rangle_I \times \begin{cases} \operatorname{Log}(\langle f \rangle_I \langle g \rangle_I) & \text{assuming GRH} \\ (\operatorname{Log}(\langle f \rangle_I \langle g \rangle_I))^t & \end{cases}$$

The inequality assuming GRH is sharp, as can be seen by taking f to be the indicator of the primes, and $g = \mathbf{1}_0$. It is also desirable to have a form of the inequality above that holds for the maximal function

$$A^* f = \sup_N |A_N f|.$$

Our second main theorem is sparse bound for A^*. The definition of a sparse bound is postponed to Definition 5.4. Remarkably, the inequality takes the same general form, although we consider a substantially larger operator.

Theorem 1.3. *For functions $f = \mathbf{1}_F$ and $g = \mathbf{1}_G$, for finite sets $F, G \subset \mathbb{Z}$, there is a sparse collection of intervals $\mathcal{S}$ so that we have*

$$\langle A^* f, g\rangle \lesssim \sum_{I \in \mathcal{S}} \langle f\rangle_I \langle g\rangle_I (\operatorname{Log}\langle f\rangle_I \langle g\rangle_I)^t |I|,$$

where we can take $t = 1$ under GRH, and otherwise we take $t = 2$.

The sparse bound is very strong, implying weighted inequalities for the maximal operator A^*. These inequalities could be further quantified, but we do not detail those consequences, as they are essentially known. See [6]. One way to see that the sparse bound is stronger is these inequalities are a corollary.

Corollary 1.4. *The maximal operator A^* satisfies these inequalities, where $t = 1$ under GRH, and $t = 2$ otherwise. First, a sparse bound with ℓ^p norms. For all $1 < p < 2$, there holds*

$$\langle A^* \mathbf{1}_F, \mathbf{1}_G\rangle \lesssim (p-1)^{-t} \sup_{\mathcal{S}} \sum_{I \in \mathcal{S}} \langle \mathbf{1}_F\rangle_{I,p} \langle \mathbf{1}_G\rangle_{I,p} |I|. \tag{1.5}$$

Second, the restricted weak-type inequalities

$$\sup_{0<\lambda<1} \frac{\lambda}{(\operatorname{Log}\lambda)^t} |\{A^* \mathbf{1}_F > \lambda\}| \lesssim |F|. \tag{1.6}$$

Third, the weak-type inequality below holds for finitely supported non-negative functions f on $\mathbb{Z}$

$$\sup_{\lambda>0} \lambda |\{A^* f > \lambda\}| \lesssim \|f\|_{\ell(\log \ell)^t (\log\log \ell)} \tag{1.7}$$

where the last norm is defined in §6.

This subject is an outgrowth of Bourgain's fundamental work on arithmetic ergodic theorems [1, 3]. These inequalities proved therein focused on the diagonal case, principally ℓ^p to ℓ^p estimates for maximal functions. Bourgain's work has been very influential, with a very rich and sophisticated theory devoted to the diagonal estimates. We point to [12, 19], and very recently [22, 24]. The subject is very rich, and the reader should consult the references in these papers.

Shortly after Bourgain's first results, Wierdl [27] studied the primes, and the simpler form of the Circle method in that case allowed him to prove diagonal inequalities for all $p > 1$, which was a novel result at that time. The result was revisited by Mirek and Trojan [20]. The unconditional version of the endpoint result (1.6) above is the main result of Trojan [25]. The approach of this paper differs in some important aspects from the one in [25]. (The low/high decomposition is dramatically different, to point to the single largest difference.)

The subject of sparse bounds originated in harmonic analysis, with a detailed set of applications in the survey [21], with a wide set of references therein. The paper [4] initiated the study of sparse bounds in the discrete setting. While the result in that paper of an 'ϵ improvement' nature, for averages it turns out there are very good results available, as was first established for the discrete sphere in [10, 14]. There is a rich theory here, with a range of inequalities for the Magyar-Stein-Wainger [17] maximal function in [15]. Nearly sharp results for certain polynomial averages are established in [5, 9], and a surprisingly good estimate for arbitrary polynomials is in [7]. The latter result plays an interesting role in the innovative result of Krause, Mirek and Tao [16].

The ℓ^p improving property for the primes was investigated in [8], but not at the endpoint. That paper result established the first weighted estimates for the averages for the prime numbers. This paper establishes the sharp results, under GRH. Mirek [18] addresses the diagonal case for Piatetski-Shapiro primes. It would be interesting to obtain ℓ^p improving estimates in this case.

Our proof uses the Circle Method to approximate the Fourier multiplier, following Bourgain [1]. In the unconditional case, we use Page's Theorem, which leads to the appearance of exceptional characters in the Circle method. Under GRH, there are no exceptional characters, and one can identify, as is well known, a very good approximation to the multiplier.

The Fourier multiplier is decomposed at the end of §3 in such a way to fit an interpolation argument of Bourgain [2], also see [11]. We call it the High/Low Frequency method. To acheive the endpoint results, this decomposition has to be carefully phrased. There are two additional features of this decomposition we found necessary to add in. First, certain difficulties associated with Ramanujan sums are addressed by making a significant change to a Low Frequency term. The sum defining the Low Frequency term (3.12) is over all Q-smooth square free denominators. Here, the integer Q can vary widely, as small as 1 and as large as $N^{1/10}$, say. (The largest Q-smooth square denominator will be of the order of e^Q.) Second, in the unconditional case, the exceptional characters are grouped into their own term. As it turns out, they can be viewed as part of the Low Frequency term. The properties we need for the High/Low method are detailed in §4. The following sections are applications of those properties.

2. Notation

We write $A \ll B$ if there is a constant C so that $A \le CB$. In such instances, the exact nature of the constant is not important.

Let $\mathcal{F}$ denote the Fourier transform on $\mathbb{R}$, defined for by

$$\mathcal{F}f(\xi) = \int_{\mathbb{R}} f(x) e^{-2\pi i x \xi} \, dx, \qquad f \in L^1(\mathbb{R}).$$

The Fourier transform on $\mathbb{Z}$ is denoted by $\widehat{f}$, defined by

$$\widehat{f}(\xi) = \sum_{n \in \mathbb{Z}} f(n) e^{-2\pi i n \xi}, \qquad f \in \ell^1(\mathbb{Z}).$$

Throughout, we denote $A_q = \{a \in \mathbb{Z}/q\mathbb{Z} : (a,q) = 1\}$, so that $|A_q| = \phi(q)$, the totient function. We have

$$\frac{q}{\operatorname{Log}\operatorname{Log} q} \ll \phi(q) \le q - 1. \tag{2.1}$$

It is known that for non-principal charactersχ, we have $|G(\chi, a)| < q^{-\frac{1}{2}}$, see [13, Chapter 3]. In particular, if χ is the principal character, then we get Ramanujan's sum

$$c_q(n) := \phi(q) G(\mathbf{1}_{A_q}, a) = \sum_{r \in A_q} e\Big(\frac{ra}{q}\Big).$$

Let χ_q denote the exceptional character. It is a non-trivial quadratic Dirichlet character modulo q, that is χ_q takes values $-1, 0, 1$, and takes the value -1 at least once. We also know that χ_q is primitive, namely that its period is q. As a matter of convenience, if q does not have an exceptional character, we will set $\chi_q \equiv 0$, and $\beta_q = 1$. These properties are important to Lemma 4.9.

Page's Theorem uses the exceptional characters to give an approximation to the prime counting function. Counting primes in an arithmetic progression of modulus q, we have

$$\psi(N;q,r) - \frac{N}{\phi(q)} + \frac{\chi_q(x)}{\phi(q)}\beta_q^{-1}x^{\beta_q} \ll Ne^{c\sqrt{\log N}}. \tag{2.2}$$

3. Approximations of the kernel

Denote the kernel of A_N with the same symbol, so that $A_N(x) = N^{-1}\sum_{n\le N}\Lambda(n)\delta_n(x)$. It follows that

$$\widehat{A_N}(\xi) = \frac{1}{N}\sum_{n\le N}\Lambda(n)e^{-2\pi n\xi}.$$

The core of the paper is the approximation to $\widehat{A_N}(\xi)$, and its further properties, detailed in the next section.

Set

$$M_N^\beta = \frac{1}{N\beta}\sum_{n\le N}[n^\beta - (n-1)^\beta]\delta_n, \qquad \tfrac{1}{2} < \beta \le 1.$$

We write $M_N = M_N^1$ when $\beta = 1$, which is the standard average. For $\beta < 1$, these are not averaging operators. They are the operators associated to the exceptional characters. The Fourier transforms are straight forward to estimate.

Proposition 3.1. *We have the estimates*

$$|\widehat{M_N}(\xi)| \ll \min\{1, (N|\xi|)^{-1}\}, \tag{3.2}$$

$$|\widehat{M_N^\beta}(\xi)| \ll (N|\xi|)^{-1}, \tag{3.3}$$

$$|\widehat{M_N^\beta}(\xi) - \beta^{-1}N^{\beta-1}| \ll N^\beta|\xi|.$$

For integers q and $a \in A_q$,

$$\widehat{L_N^{a,q}}(\xi) = G(\mathbf{1}_{A_q}, a)\widehat{M_N}(\xi) - G(\chi_q, a)\widehat{M_N^{\beta_q}}(\xi) \tag{3.4}$$

We state the approximation to the kernel at rational point, with small denominator.

Lemma 3.5. *Assume that* $|\xi - \frac{a}{q}| \le N^{-1}Q$ *for some* $1 \le a \le q \le Q$ *and* $\gcd(a,q) = 1$*. Then*

$$\widehat{A_N}(\xi) = \widehat{L_N^{a,q}}(\xi - \tfrac{a}{q}) + \begin{cases} O(QN^{-\frac{1}{2}+\epsilon}), & \textit{Assuming GRH} \\ O(Qe^{-c\sqrt{n}}), & \textit{Otherwise} \end{cases} \tag{3.6}$$

Proof. We proceed under GRH, and return to the unconditional case at the end of the argument. The key point is that we have the approximation (1.1) for $\psi(N; q, r)$. Set $\alpha := \xi - \frac{a}{q}$. Using Abel summation, we can write

$$N\widehat{M_N}(\alpha) = Ne(\alpha N) - \sqrt{N}e(\alpha\sqrt{N}) - 2\pi i\alpha \int_{\sqrt{N}}^{N} e^{t\alpha}\, dt + O(\sqrt{N}).$$

Turning to the primes, we separate out the sum below according to residue classes mod q. Since $\xi = \frac{a}{q} + \alpha$,

$$\begin{aligned} \sum_{\ell \le N} e(\xi\ell)\Lambda(\ell) &= \sum_{\substack{0 \le r \le q \\ \gcd(r,q)=1}} \sum_{\substack{\ell \le N \\ \ell \equiv r \mod q}} e(\xi\ell)\Lambda(\ell) \\ &= \sum_{r \in A_q} e\left(\tfrac{ra}{q}\right) \sum_{\substack{\ell \le N \\ \ell \equiv r \mod q}} e(\alpha\ell)\Lambda(\ell). \end{aligned}$$

Examine the inner sum. Using Abel's summation formula, and the notation ψ for prime counting function, we have

$$\begin{aligned} \sum_{\substack{\ell \le N \\ \ell \equiv r \mod q}} e(\alpha\ell)\Lambda(\ell) &= \psi(N; q, r)e(\alpha N) - \psi(\sqrt{N}; q, r)e(\alpha\sqrt{N}) \\ &\quad - 2\pi i\alpha \int_{\sqrt{N}}^{N} \psi(t; q, r)e(\alpha t)dt + O(\sqrt{N}). \end{aligned}$$

At this point we can use the Generalized Riemann Hypothesis. From (1.1), it follows that

$$\begin{aligned}
&\sum_{\substack{\ell\le N\\ \ell\equiv r \mod q}} e(\alpha\ell)\Lambda(\ell) - \frac{N}{\phi(q)}\widehat{M_N}(\alpha)\\
&\quad = (\psi(N;q,r) - \frac{N}{\phi(q)}e(\alpha N))e(\alpha N)\\
&\qquad - 2\pi i\alpha\int_{\sqrt{N}}^{N} e(t\alpha)(\psi(t;q,r)-t)\,dt + O(\sqrt{N})\\
&\quad \ll N^{\frac12+\epsilon} + \frac{Q}{N}\int_{\sqrt{N}}^{N} t^{\frac12+\epsilon}dt + O(N^{\frac12+\epsilon})\\
&\quad \ll QN^{\frac12+\epsilon}.
\end{aligned}$$

The proof without GRH uses Page's Theorem (2.2) in place of (1.1). We omit the details. □

The previous Lemma approximates $\widehat{A_N}(\xi)$ near a rational point. We extend this approximation to the entire circle. This is done with these definitions.

$$\begin{aligned}
\widehat{V_{s,n}}(\xi) &= \sum_{a/q\in\mathcal{R}_s} G(\mathbf{1}_{A_q},a)\widehat{M_N}(\xi-a/q)\eta_s(\xi-a/q),\\
\widehat{W_{s,n}}(\xi) &= \sum_{a/q\in\mathcal{R}_s} G(\chi_q,a)\widehat{M_N^{\beta_q}}(\xi-a/q)\eta_s(\xi-a/q),\\
\mathcal{R}_s &= \{a/q \ :\ a\in A_q,\ 2^s\le q<2^{s+1}\},
\end{aligned}$$

and $\mathcal{R}_0=\{0\}$. Further $\mathbf{1}_{[-1/4,1/4]}\le\eta\le\mathbf{1}_{[-1/2,1/2]}$, and $\eta_s(\xi)=\eta(4^s\xi)$. In (3.11), recall that if q is not exceptional, we have $\chi_q=0$. Otherwise, χ_q is the associated exceptional Dirichlet character. Given integer $N=2^n$, set

$$\tilde{N} = \begin{cases} e^{c\sqrt{n}/4} & \text{where } c \text{ is as in (3.6)}\\ N^{1/5} & \text{under GRH}\end{cases}$$

Lemma 3.7. *Let $N=2^n$. Write $A_N=B_N+\mathrm{Err}_N$, where*

$$B_N = \sum_{s\,:\,2^s<(\tilde{N})^{1/400}} V_{s,n}-W_{s,n}. \tag{3.8}$$

Then, we have $\|\mathrm{Err}_N f\|_{\ell^2}\ll(\tilde{N})^{-1/1000}\|f\|_{\ell^2}$.

Proof. We estimate the ℓ^2 norm by Plancherel's Theorem. That is, we bound

$$\|\widehat{A_N} - \widehat{B_N}\|_{L^\infty(\mathbb{T})} \ll (\tilde{N})^{-1/1000}.$$

Fix $\xi \in \mathbb{T}$, where we will estimate the L^∞ norm above. By Dirichlet's Theorem, there are relatively prime integers a, q with $0 \le a < q \le (\tilde{N})^{1/5}$ with

$$|\xi - a/q| < \frac{1}{q^2}.$$

The argument now splits into cases, depending upon the size of q.

Assume that $(\tilde{N})^{1/400} < q \le (\tilde{N})^{1/5}$. This is a situation for which the classical Vinogradov inequality [26, Chapter 9] was designed. That estimate is however is not enough for our purposes. Instead we use [13, Thm 13.6] for the estimate below.

$$|\widehat{A_N}(\xi)| \ll (q^{-1/2} + (q/N)^{1/2} + N^{-1/5}) \log^3 N \ll (\tilde{N})^{-1/1000}.$$

So, in this case we should also see that $\widehat{B_N}(\xi)$ satisfies the same bound. The function $\widehat{B_N}$ is a sum over $\widehat{V_{s,n}}$ and $\widehat{W_{s,n}}$. The argument for both is the same. Suppose that $\widehat{V_{s,n}}(\xi) \neq 0$. The supporting intervals for $\eta_s(\xi - a/q)$ for $a/q \in \mathcal{R}_s$ are pairwise disjoint. We must have $|\xi - a_0/q_0| < 2^{-2s}$ for some $a_0/q_0 \in \mathcal{R}_s$, where $2^s < (\tilde{N})^{1/400}$. Then,

$$|\xi - a_0/q_0| \ge |a_0/q_0 - a/q| - |\xi - a/q| \ge (qq_0)^{-1} - q^{-2} \ge q_0^{-4}.$$

But then by the decay estimate (3.2), we have

$$|G(\mathbf{1}_{A_q}, a_0)\widehat{M_N}(\xi - a_0/q_0)| \ll (Nq_0^{-4})^{-1} \ll N^{-1}(\tilde{N})^{1/100}$$

This estimate is summed over $s \le (\tilde{N})^{1/400}$ to conclude this case.

Proceed under the assumption that $q \le N_0 = (\tilde{N})^{1/400}$. From Lemma 3.5, the inequality (3.6) holds.

$$\widehat{A_N}(\xi) = \widehat{L_N^{a,q}}(\xi - \tfrac{a}{q}) + O(N_0^{-1/2})$$

The Big O term is as is claimed, so we verify that $\widehat{B_N}(\xi) - \widehat{L_N^{a,q}}(\xi - \frac{a}{q}) \ll N_0^{-1/2}$.

The analysis depends upon how close ξ is to a/q. Suppose that $|\xi - a/q| < \frac{1}{4}N_0^{-2}$. Then a/q is the unique rational b/r with $(b, r) = 1$ and $0 \le$

$b < r \leq N_0$ that meets this criteria. That means that

$$\widehat{B_N}(\xi) = \widehat{L_N^{a,q}}(\xi - a/q)\eta_s(\xi - a/q)$$

where in the last term on the right, $2^s \leq q < 2^{s+1}$. By definition $\eta_s(\xi - a/q) = \eta(4^s(\xi - a/q))$, which equals one by assumption on ξ. That completes this case.

Continuing, suppose that there is no a/q with $|\xi - a/q| < N_0^{-2}$. The point is that we have the decay estimates (3.2) and (3.3) which imply

$$|\widehat{M_N}(\xi - a/q)| + |\widehat{M_N^{\beta}}(\xi - a/q)| \ll [N(\xi - a/q)]^{-1} \ll \frac{N_0^2}{N} \ll N^{-3/5}.$$

But then, from the definition (3.4), we have

$$|\widehat{L_N^{a,q}}(\xi - \tfrac{a}{q})| \ll N^{-1/5}.$$

And as well, trivially bounding Gauss sums by 1, we have

$$|\widehat{B_N}(\xi)| \ll \frac{n^{3/5}}{N} \ll N^{-1/5},$$

by just summing over all $a/q \in \mathcal{R}_s$, with $s < (\tilde{N})^{1/400}$. That completes the proof. □

The discussion to this point is of a standard nature. We state here a decomposition of the operator B_N defined in (3.8). It encodes our High/Low/Exceptional decomposition, and requires some care to phrase, in order to prove our endpoint type results for the prime averages. It depends upon a supplementary parameter Q. This parameter Q will play two roles, controlling the size and smoothness of denominators. Recall that an integer q is *Q-smooth* if all of its prime factors are less than Q. Let $\mathbb{S}_Q$ be the collection of square-free Q-smooth integers.

$$\widehat{V_{s,n}^{Q,\mathrm{lo}}}(\xi) = \sum_{\substack{a/q \in \mathcal{R}_s \\ q \in \mathbb{S}_Q}} G(\mathbf{1}_{A_q}, a)\widehat{M_N}(\xi - a/q)\eta_s(\xi - a/q), \tag{3.9}$$

$$\widehat{V_{s,n}^{Q,\mathrm{hi}}}(\xi) = \sum_{\substack{a/q \in \mathcal{R}_s \\ q \notin \mathbb{S}_Q}} G(\mathbf{1}_{A_q}, a)\widehat{M_N}(\xi - a/q)\eta_s(\xi - a/q), \tag{3.10}$$

$$\widehat{W_{s,n}}(\xi) = \sum_{a/q \in \mathcal{R}_s} G(\chi_q, a)\widehat{M_N^{\beta_q}}(\xi - a/q)\eta_s(\xi - a/q), \tag{3.11}$$

Define

$$\mathrm{Lo}_{Q,N} = \sum_s V_{s,n}^{Q,\mathrm{lo}}, \tag{3.12}$$

$$\mathrm{Hi}_{Q,N} = \sum_{s\,:\,Q \le 2^s \le (\tilde{N})^{1/400}} V_{s,n}^{Q,\mathrm{hi}} - W_{s,n} \tag{3.13}$$

$$\mathrm{Ex}_{Q,N} = \sum_{s\,:\,2^s \le Q} W_{s,n} \tag{3.14}$$

Concerning these definitions, in the Low term (3.12), there is no restriction on s, but the sum only depends upon the finite number of square-free Q-smooth numbers in $\mathbb{S}_Q$. (Due to (4.8), the non-square free integers will not contribute to the sum.) The largest integer in $\mathbb{S}_Q$ will be about e^Q, and the value of Q can be as big as $\tilde{N}$. In the High term (3.13), there are two parts associated with the principal and exceptional characters. For the principal characters, we exclude the square free Q-smooth denominators which are both larger than Q and less than $(\tilde{N})^{1/400}$. These are included in the Low term. We include all the denominators for the exceptional characters. In the Exceptional term (3.14), we just impose the restriction on the size of the denominator to be not more than Q. This will be part of the Low term.

The sum of these three terms well approximates B_N.

Proposition 3.15. *Let $1 \le Q \le \tilde{N}$. We have the estimate*

$$\|\mathrm{Err}'_{Q,N} f\|_{\ell^2} \lesssim (\tilde{N})^{-1/2} \|f\|_{\ell^2}, \tag{3.16}$$

where

$$\mathrm{Err}'_{Q,N} = \mathrm{Lo}_{Q,N} + \mathrm{Hi}_{Q,N} + \mathrm{Ex}_N + \mathrm{Err}_N - B_N. \tag{3.17}$$

Proof. From (3.8), we see that

$$\widehat{\mathrm{Err}'_N}(\xi) = \sum_{s\,:\,2^s > (\tilde{N})^{1/400}} \widehat{V_{s,n}^{Q,\mathrm{lo}}}(\xi)$$

Recalling the definition of $V_{s,n}^{Q,\mathrm{lo}}$ from (3.9), it is straight forward to estimate this last sum in $L^\infty(\mathbb{T})$, using the Gauss sum estimate $G(\mathbf{1}_{A_q}, a) \ll \frac{\mathrm{Log}\,\mathrm{Log}\, q}{q}$. □

4. Properties of the high, low and exceptional terms

The further properties of the High, Low and Exceptional terms are given here, in that order.

4.1. The high terms

We have the ℓ^2 estimates for the fixed scale, and and for the supremum over large scales, for the High term defined in (3.13). Note that the supremum is larger by a logarithmic factor.

Lemma 4.1. *We have the inequalities*

$$\|\mathrm{Hi}_{Q,N}\|_{\ell^2\to\ell^2} \lesssim \frac{\log\log Q}{Q}, \tag{4.2}$$

$$\|\sup_{N>Q^2} |\mathrm{Hi}_{Q,N}\, f|\|_2 \lesssim \frac{\log\log Q\cdot\log Q}{Q}\|f\|_{\ell^2}. \tag{4.3}$$

We comment that the insertion of the Q smooth property into the definition of $V_{s,n}^{Q,\mathrm{hi}}$ in (3.10) is immaterial to this argument.

Proof. Below, we assume that there are no exceptional characters, as a matter of convenience as the exceptional characters are treated in exactly the same manner. For the inequality (4.2), we have from the definition of the High term in (3.13), and (3.10),

$$\begin{aligned}
\|\mathrm{Hi}_{Q,N}\|_{\ell^2\to\ell^2} &= \|\widehat{\mathrm{Hi}_{Q,N}}\|_{L^\infty(\mathbb{T})} \\
&= \Big\| \sum_{s\,:\,Q\le 2^s\le \tilde N} \widehat{V_{s,n}^{Q,\mathrm{hi}}} \Big\|_{L^\infty(\mathbb{T})} \\
&\le \sum_{s\,:\,Q\le 2^s\le \tilde N} \|\widehat{V_{s,n}^{Q,\mathrm{hi}}}\|_{L^\infty(\mathbb{T})} \\
&\le \sum_{s\,:\,Q\le 2^s\le \tilde N} \max_{2^s\le q<2^{s+1}} \max_{a\in A_q} |G(\mathbf{1}_{A_q}, a)| \\
&\ll \sum_{s\,:\,Q\le 2^s\le \tilde N} \max_{2^s\le q<2^{s+1}} \frac{1}{\phi(q)} \\
&\ll \sum_{s\,:\,Q\le 2^s} \log s\cdot 2^{-s} \ll \frac{\log\log Q}{Q}.
\end{aligned}$$

The first line is Plancherel, and the subsequent lines depend upon definitions, and the fact that the functions below are disjointly supported.

$$\{\eta_s(\cdot - a/q) \ : \ 2^s \le q < 2^{s+1}, \ a \in A_q\}.$$

Last of all, we use a well known lower bound $\phi(q) \gg q/\log\log q$.

For the maximal inequality (4.3), we have an additional logarithmic term. This is direct consequence of the Bourgain multi-frequency inequality, stated in Lemma 4.4. We then have

$$\begin{aligned}\| \sup_{N>Q^2} |\mathrm{Hi}_{Q,N}\, f|\|_{\ell^2} &\le \sum_{s\,:\,Q\le 2^s} \Big\| \sup_{N>Q^2} |V^{Q,\mathrm{hi}}_{s,n} f|\Big\|_{\ell^2} \\ &\ll \sum_{s\,:\,Q\le 2^s} s \cdot \max_{2^s\le q<2^{s+1}} \frac{1}{\phi(q)} \cdot \|f\|_{\ell^2} \\ &\lesssim \frac{\log Q \cdot \log\log Q}{Q} \|f\|_{\ell^2}.\end{aligned}$$

□

Lemma 4.4. *Let $\theta_1, \dots, \theta_J$ be points in $\mathbb{T}$ with $\min_{j\neq k}|\theta_j - \theta_k| > 2^{-2s_0+2}$. We have the inequality*

$$\Big\| \sup_{N>4^{s_0}} \Big| \sum_{j=1}^{J} \mathcal{F}^{-1}\Big(\widehat{f} \sum_{j=1}^{J} \tilde{M}_N(\cdot - \theta_j)\eta_{s_0}(\cdot - a/q)\Big)\Big|\Big\|_{\ell^2} \ll \log J \cdot \|f\|_{\ell^2}.$$

This is one of the main results of [3]. It is stated therein with a higher power of $\log J$. But it is well known that the inequality holds with a single power of $\log J$. This is discussed in detail in [8].

4.2. The low terms

From the Low terms defined in (3.12), the property is

Lemma 4.5. *For a functions f, g supported on interval I of length $N = 2^n$, we have*

$$N^{-1}\langle \mathrm{Lo}_{Q,N} * f, g\rangle \ll \log Q \cdot \langle f\rangle_I \langle g\rangle_I. \tag{4.6}$$

The following Möbius Lemma is well known.

Lemma 4.7. *For each q, we have*

$$\sum_{a\in A_q} G(\mathbf{1}_{A_q}, a)\mathcal{F}^{-1}(\widehat{M}_N \cdot \eta_s(\cdot - a/q))(x) = \frac{\mu(q)}{\phi(q)} c_q(-x). \tag{4.8}$$

Proof. Compute

$$\begin{aligned}
&\sum_{a\in A_q} G(\mathbf{1}_{A_q}, a)\mathcal{F}^{-1}(\widehat{M}_N \cdot \eta_s(\cdot - a/q))(x) \\
&\qquad = M_N * \mathcal{F}^{-1}\eta_s(x) \sum_{a\in A_q} G(\mathbf{1}_{A_q}, a) e(ax/q).
\end{aligned}$$

We focus on the last sum above, namely

$$\begin{aligned}
S_q(x) &= \sum_{a\in A_q} G(\mathbf{1}_q, a) e(xa/q) \\
&= \frac{1}{\phi(q)} \sum_{r\in A_q} \sum_{a\in A_q} e(a(r+x)/q) \\
&= \frac{1}{\phi(q)} \sum_{r\in A_q} c_q(r+x) = \frac{\mu(q)}{\phi(q)} c_q(-x).
\end{aligned}$$

The last line uses Cohen's identity. □

The two steps of inserting of the property of being Q smooth in (3.9), as well as dropping an restriction on s in (3.12), were made for this proof.

Proof of Lemma 4.5. By (4.8), the kernel of the operator $\mathrm{Lo}_{Q,N}$ is

$$\begin{aligned}
\mathrm{Lo}_{Q,N}(x) &= M_N * \mathcal{F}^{-1}\eta_s(x) \cdot S(-x), \\
\text{where} \quad S(x) &= \sum_{q\in \mathbb{S}_Q} \frac{\mu(q)}{\phi(q)} c_q(x).
\end{aligned}$$

We establish a pointwise bound $\|S\|_{\ell^\infty} \ll \log Q$, which proves the Lemma.

Assume $x \neq 0$. We exploit the multiplicative properties of the summands, as well as the fact that if prime p divides x, we have $\frac{\mu_p(x)}{\phi(p)} c_q(x) = \mu_p(x)$. Let $\mathcal{Q}_1$ be the primes $p < Q$ such that $(p, x) = 1$, and set $\mathcal{Q}_2$ to be the primes less than Q which are not in $\mathcal{Q}_1$.

The multiplicative aspect of the sums allows us to write

$$\frac{\mu(q)}{\phi(q)} c_q(-x) = \frac{\mu(q_1)}{\phi(q_1)} c_{q_1}(-x) \cdot \mu(q_2)$$

where $q = q_1 q_2$, and all prime factors of q_j are in $\mathcal{Q}_j$. If $\mathcal{Q}_j$ is empty, set $q_j = 1$. Thus, $S(x) = S_1(x) S_2(x)$, where the two terms are associated with $\mathcal{Q}_1$ and $\mathcal{Q}_2$ respectively. We have

$$\begin{aligned} S_1(x) &= \sum_{q \text{ is } \mathcal{Q}_1 \text{ smooth}} \frac{\mu(q)}{\phi(q)} c_q(-x) \\ &= \prod_{p \in \mathcal{Q}_1} 1 + \frac{\mu(p) c_p(-x)}{\phi(p)} \\ &= \prod_{p \in \mathcal{Q}_1} 1 + \frac{1}{p-1} = A_x. \end{aligned}$$

This is so, since $\mu(p) c_p(x) = 1$. It is a straight forward consequence of the Prime Number Theorem that $A_x \ll \log Q$. Here, and below, we say that q is $\mathcal{Q}$ smooth if all the prime factors of q are in the set of primes $\mathcal{Q}$.

The second term is as below, where $d = |\mathcal{Q}_2|$. Here, in the definition (3.12), there is no restriction on s, hence all the smooth square free numbers are included. If $\mathcal{Q}_2 = \emptyset$, then $S_2(x) = 1$, otherwise

$$\begin{aligned} S_2(x) &= \sum_{q \text{ is } \mathcal{Q}_2 \text{ smooth}} \mu(q) \\ &= \sum_{j=1}^{d} \binom{d}{j} (-1)^j \\ &= -1 + \sum_{j=0}^{d} \binom{d}{j} (-1)^j = -1. \end{aligned}$$

If $x = 0$, then $S(0) = S_2(x) = -1$. That completes the proof. □

4.3. The exceptional term

The Exceptional terms are always of a smaller order than the Low terms.

Lemma 4.9. *Let χ be an exceptional character modulo q. For $x \in \mathbb{Z}$,*

$$\Big|\sum_{a\in A_q} G(\chi,a)e(xa/q)\Big| = \frac{q}{\phi(q)} \tag{4.10}$$

provided $(x,q)=1$, otherwise the sum is zero.

Proof. It is also known that exceptional characters are primitive - see [13, Theorem 5.27]. So the sum is zero if $(x,q)>1$. We use the common notation

$$\tau(\chi,x) = \sum_{a\in A_q} \chi(a)e(ax/q)$$

which is $\phi(q)G(\chi,x)$. Assuming $(x,q)=1$,

$$\tau(\chi,a) = \tau(\chi,1).$$

This leads immediately to

$$\begin{aligned}\sum_{a\in A_q} \tau(\chi,a)e(\frac{ax}{q}) &= \tau(\chi,1)\sum_{a\in A_q}\chi(a)e(-\frac{ax}{q})\\ &= \frac{\tau(\chi)\overline{\tau(\chi,x)}}{\phi(q)} = \frac{|\tau(\chi)|^2\overline{\chi(x)}}{\phi(q)}.\end{aligned}$$

It is known that $|\tau(\chi)|^2 = q$ for primitive characters. And the exceptional character is quadratic, so this completes the proof. □

Lemma 4.11. *For a function f supported on interval I of length $N=2^n$, we have*

$$\langle \mathrm{Ex}_{Q,N} * f\rangle_\infty \ll (\log\log Q)^2 \cdot \langle f\rangle_I. \tag{4.12}$$

The term on the left is defined in (3.14).

Proof. Following the argument from Lemma 4.5, we have

$$\mathrm{Ex}_{Q,N}(x) = \sum_{q<Q}\sum_{a\in A_q} G(\chi_q,a)e(xa/q)\cdot M_N^{\beta_v} * \mathcal{F}^{-1}\eta_{s_q}(x).$$

Above, $2^{s_q} \le q < 2^{s_q+1}$. The interior sum above is estimated in (4.10). Using the lower bound on the totient function in (2.1), we have

$$\mathrm{Ex}_{Q,N}(x) f \ll \log\log Q \cdot \langle f \rangle_I \sum_{\substack{q<Q \\ q \text{ exceptional}}} 1.$$

We know that the exceptional q grow at the rate of a double exponential, that is for q_v being the vth exceptional q, we have $q_v \gg C^{C^v}$, for some $C > 1$. It follows that the sum above is at most $\log\log Q$. □

5. Proofs of the fixed scale and sparse bounds

Proof of Theorem 1.2. Let $N = 2^n$, and recall that $f = \mathbf{1}_F$ and $g = \mathbf{1}_G$ where $F, G \subset I$, and interval of length N.

Let us address the case in which we do not assume GRH. We always have the estimate

$$N^{-1}\langle A_N f, g\rangle \lesssim n \cdot \langle f \rangle_I \langle g \rangle_I. \tag{5.1}$$

Hence, if we have $\langle f \rangle_I \langle g \rangle_I \ll e^{-c\sqrt{n}/100}$, the inequality with a squared log follows.

We assume that $e^{-c\sqrt{n}} \ll \langle f \rangle_I \langle g \rangle_I$, and then prove a better estimate. We turn to the Low/High/Exceptional decomposition in (3.12)–(3.14), for a choice of integer Q that we will specify. We have

$$A_N = \mathrm{Lo}_{Q,N} + \mathrm{Hi}_{Q,N} - \mathrm{Ex}_{Q,N} + \mathrm{Err}_N + \mathrm{Err}'_N \tag{5.2}$$

These terms are defined (3.12), (3.13), (3.14), (3.8) and (3.17) respectively.

For the 'High' term we have by (4.2),

$$N^{-1}|\langle \mathrm{Hi}_{Q,N}\, f, g\rangle| \lesssim \frac{\log\log Q}{Q} \langle f \rangle_{I,2} \langle g \rangle_{I,2}$$

The same inequality holds for both $\mathrm{Err}_{Q,N}\, f$ and $\mathrm{Err}'_{Q,N}\, f$ by Lemma 3.7 and Proposition 3.15.

Concering the Low term, by (4.6), we have

$$N^{-1}|\langle \mathrm{Lo}_{Q,N}\, f, g\rangle| \lesssim \log Q \langle f \rangle_I \langle g \rangle_I$$

The Exceptional term satisfies the same estimate by (4.12).

Combining estimates, choose Q to minimize the right hand side, namely

$$(5.3) \qquad N^{-1}\langle A_N f, g\rangle \lesssim \frac{\log\log Q}{Q}\big[\langle f\rangle_I \langle g\rangle_I\big]^{1/2} + \log Q \cdot \langle f\rangle_I \langle g\rangle_I .$$

This value of Q is

$$Q\frac{\log Q}{\log\log Q} \simeq \big[\langle f\rangle_I \langle g\rangle_I\big]^{-1/2}.$$

Since $e^{-c\sqrt{n}} \ll \langle f\rangle_I \langle g\rangle_I$, this is an allowed choice of Q. And, then, we prove the desired inequality, but only need a single power of logarithm.

Assuming GRH, from (5.1), we see that the inequality to prove is always true provided $\langle f\rangle_I \langle g\rangle_I < cN^{-1/4}$. Assuming this inequality fails, we follow the same line of reasoning above that leads to (5.3). That value of Q will be at most $N^{1/4}$, so the proof will complete, to show the bound with a single power of the logarithmic term. □

Turning to the sparse bounds, let us begin with the definitions.

Definition 5.4. *A collection of intervals $\mathcal{S}$ is called* sparse *if to each interval $I \in \mathcal{S}$, there is a set $E_I \subset I$ so that $4|E_I| \geq |I|$ and the collection $\{E_I : I \in \mathcal{S}\}$ are pairwise disjoint. All intervals will be finite sets of consecutive integers in $\mathbb{Z}$.*

The form of the sparse bound in Theorem 1.3 strongly suggests that one use a recursive method of proof. (Which is indeed the common method.) To formalize it, we start with the notion of a *linearized* maximal function. Namely, to bound the maximal function $A^* f$, it suffices to bound $A_{\tau(x)} f(x)$, where $\tau : \mathbb{Z} \to \{2^n : n \in \mathbb{N}\}$ is a function, taken to realize the supremum. The supremum in the definition of $A^* f$ is always attained if f is finitely supported.

Definition 5.5. *Let I_0 an interval, and let f be supported on $3I_0$. A map $\tau : I_0 \to \{1, 2, 4, \dots, |I_0|\}$ is said to be* admissible *if*

$$\sup_{N \geq \tau(x)} M_N f(x) \leq 10\langle f\rangle_{3I_0,1}.$$

That is, τ is admissible if at all locations x, the averages of f over scales larger than $\tau(x)$ are controlled by the global average of f.

Lemma 5.6. *Let f and τ be as in Definition 5.5. Further assume that f and g are indicator functions, with g supported on I_0. Then, we have*

$$|I_0|^{-1}\langle A_\tau f, g\rangle \lesssim \langle f\rangle_{I_0,1}\langle g\rangle_{I_0,1}\cdot(\operatorname{Log}\langle f\rangle_{3I_0,1}\langle g\rangle_{I_0,1})^t,$$

where $t=1$ assuming RH, and $t=2$ otherwise.

Proof. We restrict τ to take values $1,2,4,\ldots,2^t,\ldots,$. Let $|I_0|=N_0=2^{n_0}$. We always have the inequalities

$$\begin{aligned}|I_0|^{-1}\langle A_\tau f, g\rangle &\lesssim n_0\langle f\rangle_{I_0,1}\langle g\rangle_{I_0,1}\\ |I_0|^{-1}\langle \mathbf{1}_{\tau<T}A_\tau f, g\rangle &\lesssim (\log T)\langle f\rangle_{I_0,1}\langle g\rangle_{I_0,1}.\end{aligned}$$

The top line follows from admissibility.

We begin by not assuming GRH. Then, the conclusion of the Lemma is immediate if we have $(\operatorname{Log}\langle f\rangle_{I_0,1}\langle g\rangle_{I_0,1})^2\gg n_0$. It is also immediate if $\log\tau\ll(\operatorname{Log}\langle f\rangle_{I_0,1}\langle g\rangle_{I_0,1})^2$. We proceed assuming

$$p_0^2 = C(\operatorname{Log}\langle f\rangle_{I_0,1}\langle g\rangle_{I_0,1})^2 \le c_0\min\{n_0,\log\tau\}, \tag{5.7}$$

where $0<c_0<1$ is sufficiently small.

We use the definitions in (3.12)–(3.14) for a value of $Q<e^{c\sqrt{n_0}}$ that we will specify. We address the High, Low, Exceptional and both Error terms, as in (5.2). First, the Error terms. The error terms come in the form of Err_N from Lemma 3.7 and Err'_N from (3.16). Both are similar. Concerning the second error term, From the estimate (3.17) and (5.7), we have by a straight forward square function argument,

$$\begin{aligned}\|\operatorname{Err}_{Q,\tau} f\|_2^2 &\le \sum_{n\,:\,p_0^2\le n\le n_0}\|\operatorname{Err}_{Q,2^n} f\|_{\ell^2}^2\\ &\lesssim \|f\|_{\ell^2}^2\sum_{n\,:\,p_0^2\le n\le n_0} e^{-c\sqrt{n}}\\ &\lesssim \|f\|_{\ell^2}^2\cdot p_0^2 e^{-cp_0}\lesssim \|f\|_{\ell^2}^2\cdot\langle f\rangle_{3I_0,1}\langle g\rangle_{I_0,1}.\end{aligned}$$

This provided C in (5.7) is large enough. This is a much smaller estimate than we need. The second error term in Proposition 3.15 is addressed by the same square function argument.

For the High term, apply (4.3) to see that

$$\| \sup_{N>Q^2} |\mathrm{Hi}_{Q,N}\, f| \|_2 \lesssim \frac{\log Q \cdot \log\log Q}{Q} \|f\|_{\ell^2}.$$

For the Low term the definition of admissibility and (4.6) that

$$|I_0|^{-1} |\langle \mathrm{Lo}_{Q,\tau(x)}\, f(x), g\rangle \ll (\log Q) \langle f\rangle_I \langle g\rangle_I.$$

The Exceptional term also satisfies this bound.

We conclude that

$$|I_0|^{-1} \langle A_\tau f, g\rangle \lesssim \frac{\log Q \cdot \log\log Q}{Q} \langle f\rangle_{I,2} \langle g\rangle_{I,2} + \log Q \cdot \langle f\rangle_I \langle g\rangle_I.$$

This is optimized by taking Q so that

$$\frac{Q}{\log\log Q} \simeq \left[\langle f\rangle_I \langle g\rangle_I\right]^{-1/2}.$$

And this will be an allowed value of Q since (5.7) holds. Again, the resulting estimate is better by power of the logarithmic term than what is claimed.

Under RH, the proof is very similar, but a wider range of Q's are allowed. In particular, only a single power of logarithm is needed. □

6. Proof of Corollary 1.4

The inequality (1.5) follows from the elementary identity that for $0 < x < 1$, we have

$$x(\operatorname{Log} x)^t \ll \min_{1<p<2} \frac{x}{(p-1)^t}.$$

We remark that we do not know an efficient way to pass from the restricted weak type sparse bound we have established to the similar sparse bounds for functions. The methods to do this for *norm estimates* is of course very well studied.

Proof of (1.6). There is a different inequality that is a natural consequence of the sparse bound, namely

$$\sup_\lambda \lambda \frac{|\{A^* \mathbf{1}_F > \lambda\}|}{(\operatorname{Log}|\{A^* \mathbf{1}_F > \lambda\}| \cdot |F|^{-1})} \lesssim |F|. \tag{6.1}$$

Indeed, if (1.6) were to fail, with a sufficiently large constant, it would contradict the inequality above.

Let $|G| > |F|$. We show that there is a subset $G' \subset G$, with $4|G'| \geq |G|$ with

$$\langle A^* f, \mathbf{1}_{G'} \rangle \ll |F| (\operatorname{Log} |F|/|G|)^t \tag{6.2}$$

This implies (6.1) by taking $G = \{A^* f > \lambda\}$, for $0 < \lambda < 1$.

In the opposite case, take G' to be

$$G' = G \setminus \{Mf > K\rho\}, \qquad \rho = |F| \cdot |G|^{-1}$$

where M is the ordinary maximal function. By the usual weak ℓ^1 inequality for M, for K sufficiently large, we have $4|G'| > |G|$. Let $g = \mathbf{1}_{G'}$. Apply the sparse bound for A^* to see that

$$\langle A^* f, g \rangle \ll \sum_{I \in \mathcal{S}} \langle f \rangle_I \langle g \rangle_I (\operatorname{Log} \langle f \rangle_I \langle g \rangle_I)^t |I|.$$

We can assume that for all intervals $I \in \mathcal{S}$, that we have $\langle g \rangle_I > 0$. That means that $\langle f \rangle_I \leq K|F|/|G|$. Turn to a pigeonhole argument. Divide the collection $\mathcal{S}$ into subcollections $\bigcup_{j,k \geq 0} \mathcal{S}_{j,k}$ where

$$\mathcal{S}_{j,k} = \{I \in \mathcal{S} \ : \ 2^{-j-1} K\rho < \langle f \rangle_I \leq 2^{-j} K\rho, \ 2^{-k-1} < \langle g \rangle_I \leq 2^{-k}\}.$$

Then, we have

$$\begin{aligned}
\langle A^* f, g \rangle &\ll \sum_{j,k \geq 0} \sum_{I \in \mathcal{S}_{j,k}} \langle f \rangle_I \langle g \rangle_I (\operatorname{Log} \langle f \rangle_I \langle g \rangle_I)^t |I| \\
&\ll |F| \cdot |G|^{-1} \sum_{j,k \geq 0} 2^{-j-k} (j + k + \operatorname{Log} \rho)^t \sum_{I \in \mathcal{S}_{j,k}} |I| \\
&\ll |F| \cdot |G|^{-1} \sum_{j,k \geq 0} 2^{-j-k} (j + k + \operatorname{Log} \rho)^t \min\{|G| 2^j, 2^k |G|\} \\
&\ll |F| \sum_{j,k \geq 0} 2^{-j-k} (j + k + \operatorname{Log} \rho) 2^{(j+k)/2} \ll |F|.
\end{aligned}$$

Here, we have used the standard weak-type inequality for the maximal function, and the basic property of sparseness, namely

$$\sum_{I \in \mathcal{S}} |I| \lesssim \Big| \bigcup_{I \in \mathcal{S}} I \Big|.$$

This completes the proof of (6.2). $\square$

For the proof of (1.7), we need to recall the definition of the Orlicz norm. Given f finitely supported on $\mathbb{Z}$, let $f^* : [0,\infty) \to \mathbb{N}$ be the decreasing rearrangement of f. That is,

$$f^*(\lambda) = |\{x \in \mathbb{Z} \,:\, |f(x)| \geq \lambda\}|.$$

For a slowly varying function $\varphi : [0,\infty) \to [0,\infty)$, set

$$\begin{aligned} \|f\|_{\ell\varphi(\ell)} &= \int_0^\infty f^*(\lambda)\varphi(\lambda)\, d\lambda \\ &\simeq \sum_{j\in\mathbb{Z}} 2^j \varphi(2^j) f^*(2^j). \end{aligned}$$

For $\varphi(x) = 1$, this is comparable to the usual ℓ^1 estimate. For $f = \mathbf{1}_F$, note that

$$\|f\|_{\ell\varphi(\ell)} = \int_0^{|F|} \varphi(\lambda)\, d\lambda \simeq |F|\varphi(|F|)$$

We are interested in $\varphi(x) = (\operatorname{Log} x) \cdot \operatorname{Log}\operatorname{Log} x)^t$, for $t = 1, 2$. The proof of the orlicz norm estimate (1.7) is below.

Proof of (1.7)*.* This argument goes back to at least [23]. Assume that the weak-type estimate for indicators (1.6) holds. Let $f \in \ell(\log \ell)^t(\log\log \ell)$ be a non-negative function of norm one. Set

$$B_j = \{x \,:\, 2^j \leq f(x) < 2^{j+1}\},$$

and set $b_j = f^*(2^j)$. We have

$$\sum_{j\leq 0} 2^j \mathbf{1}_{B_j} \leq f \leq 2 \sum_{j\leq 0} 2^j \mathbf{1}_{B_j}.$$

And, by logarithmic subadditivity for the weak-type norm, and (1.6),

$$\begin{aligned} \|A^* f\|_{1,\infty} &\ll \sum_{j\leq 0} \log(1-j) \cdot 2^j \|A^* \mathbf{1}_{B_j}\|_{1,\infty} \\ &\ll \sum_{j\leq 0} \log(1-j) \cdot 2^j |B_j| (\log |B_j|)^t \\ &\ll \sum_{j\leq 0} \log(1-j) \cdot j^t 2^j |B_j| \ll \|f\|_{\ell(\log\ell)^t(\log\log\ell)} = 1. \end{aligned}$$

Above, we appealed to $|B_j| \leq 2^{-j}$, for otherwise the norm of f is more than one. □

Acknowledgements

One of us (MTL) is a 2020 Simons Fellow and their research is supported in part by the Australian Research Council (ARC) through the research grant DP170101060. Research of all authors supported in part by grant from the US National Science Foundation, DMS-1949206.

References

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

[2] Jean Bourgain, *Estimations de certaines fonctions maximales*, C. R. Acad. Sci. Paris Sér. I Math. **301** (1985), no. 10, 499–502.

[3] ———, *Pointwise ergodic theorems for arithmetic sets*, Inst. Hautes Études Sci. Publ. Math. **69** (1989), 5–45. With an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein.

[4] Amalia Culiuc, Robert Kesler, and Michael T. Lacey, *Sparse bounds for the discrete cubic Hilbert transform*, Anal. PDE **12** (2019), no. 5, 1259–1272.

[5] Shival Dasu, Ciprian Demeter, and Bartosz Langowski, *Sharp l^p-improving estimates for the discrete paraboloid*, arXiv e-prints (February 2020), `arXiv:2002.11758`.

[6] Dorothee Frey and Bas Nieraeth, *Weak and strong type $A_1 - A_\infty$ estimates for sparsely dominated operators*, J. Geom. Anal. **29** (2019), no. 1, 247–282.

[7] Rui Han, Vjekoslav Kovač, Michael T. Lacey, José Madrid, and Fan Yang, *Improving estimates for discrete polynomial averages*, J. Fourier Anal. Appl. **26** (2020), no. 3, Paper No. 42, 11.

[8] Rui Han, Ben Krause, Michael T. Lacey, and Fan Yang, *Averages along the primes: improving and sparse bounds*, Concr. Oper. **7** (2020), no. 1, 45–54.

[9] Rui Han, Michael T Lacey, and Fan Yang, *Averages along the Square Integers: ℓ^p improving and Sparse Inequalities*, arXiv e-prints (July 2019), `arXiv:1907.05734`.

[10] Kevin Hughes, *ℓ^p-improving for discrete spherical averages*, Ann. H. Lebesgue **3** (2020), 959–980 (English, with English and French summaries).

[11] Alexandru D. Ionescu, *An endpoint estimate for the discrete spherical maximal function*, Proc. Amer. Math. Soc. **132** (2004), no. 5, 1411–1417.

[12] Alexandru D. Ionescu and Stephen Wainger, *L^p boundedness of discrete singular Radon transforms*, J. Amer. Math. Soc. **19** (2006), no. 2, 357–383.

[13] Henryk Iwaniec and Emmanuel Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, vol. 53, American Mathematical Society, Providence, RI, 2004.

[14] R. Kesler and M. T. Lacey, *ℓ^p-Improving Inequalities for Discrete Spherical Averages*, Anal. Math. **46** (2020), no. 1, 85–95.

[15] Robert Kesler, Michael T. Lacey, and Darío Mena, *Sparse bounds for the discrete spherical maximal functions*, Pure Appl. Anal. **2** (2020), no. 1, 75–92.

[16] Ben Krause, Mariusz Mirek, and Terence Tao, *Pointwise ergodic theorems for non-conventional bilinear polynomial averages*, arXiv e-prints (August 2020), `arXiv:2008.00857`.

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

[18] Mariusz Mirek, *Roth's theorem in the Piatetski-Shapiro primes*, Rev. Mat. Iberoam. **31** (2015), no. 2, 617–656.

[19] ———, *Square function estimates for discrete Radon transforms*, Anal. PDE **11** (2018), no. 3, 583–608.

[20] Mariusz Mirek and Bartosz Trojan, *Cotlar's ergodic theorem along the prime numbers*, J. Fourier Anal. Appl. **21** (2015), no. 4, 822–848.

[21] María Cristina Pereyra, *Dyadic harmonic analysis and weighted inequalities: the sparse revolution*, arXiv e-prints (December 2018), `arXiv:1812.00850`.

[22] Lillian B. Pierce, *On superorthogonality*, arXiv e-prints, `arXiv:2007.10249`.

[23] Per Sjölin, *An inequality of Paley and convergence a.e. of Walsh-Fourier series*, Ark. Mat. **7** (1969), 551–570.

[24] Terence Tao, *The Ionescu–Wainger multiplier theorem and the adeles*, arXiv e-prints (August 2020), arXiv:2008.05066.

[25] Bartosz Trojan, *Endpoint estimates for the maximal function over prime numbers*, J. Fourier Anal. Appl. **25** (2019), no. 6, 3123–3153.

[26] I. M. Vinogradov, *Elements of number theory*, Dover Publications, Inc., New York, 1954. Translated by S. Kravetz.

[27] Máté Wierdl, *Pointwise ergodic theorem along the prime numbers*, Israel J. Math. **64** (1988), no. 3, 315–336 (1989).

School of Mathematics, Georgia Institute of Technology
Atlanta GA 30332, USA
E-mail address: lacey@math.gatech.edu
E-mail address: hmousavi6@gatech.edu
E-mail address: yaghoub.rahimi@gatech.edu

Received January 25, 2021
Accepted June 10, 2021