Annals of Mathematics, **161** (2005), 1609–1636

Roth's theorem in the primes

By Ben Green*

Abstract

We show that any set containing a positive proportion of the primes contains a 3-term arithmetic progression. An important ingredient is a proof that the primes enjoy the so-called Hardy-Littlewood majorant property. We derive this by giving a new proof of a rather more general result of Bourgain which, because of a close analogy with a classical argument of Tomas and Stein from Euclidean harmonic analysis, might be called a *restriction theorem for the primes.*

1. Introduction

Arguably the second most famous result of Klaus Roth is his 1953 upper bound [21] on $r_3(N)$, defined 17 years previously by Erdős and Turán to be the cardinality of the largest set $A \subseteq [N]$ containing no nontrivial 3-term arithmetic progression (3AP). Roth was the first person to show that $r_3(N) = o(N)$. In fact, he proved the following quantitative version of this statement.

Proposition 1.1 (Roth). $r_3(N) \ll N/\log\log N$.

There was no improvement on this bound for nearly 40 years, until Heath-Brown [15] and Szemerédi [22] proved that $r_3 \ll N(\log N)^{-c}$ for some small positive constant c. Recently Bourgain [6] provided the best bound currently known.

Proposition 1.2 (Bourgain). $r_3(N) \ll N\,(\log\log N/\log N)^{1/2}$.

*The author is supported by a Fellowship of Trinity College, and for some of the period during which this work was carried out enjoyed the hospitality of Microsoft Research, Redmond WA and the Alfréd Rényi Institute of the Hungarian Academy of Sciences, Budapest. He was supported by the *Mathematics in Information Society* project carried out by Rényi Institute, in the framework of the European Community's *Confirming the International Rôle of Community Research* programme.

The methods of Heath-Brown, Szemerédi and Bourgain may be regarded as (highly nontrivial) refinements of Roth's technique. There is a feeling that Proposition 1.2 is close to the natural limit of this method. This is irritating, because the sequence of primes is not covered by these results. However it *is* known that the primes contain infinitely many 3APs.[1]

PROPOSITION 1.3 (Van der Corput). *The primes contain infinitely many* 3AP*s.*

Van der Corput's method is very similar to that used by Vinogradov to show that every large odd number is the sum of three primes. Let us also mention a paper of Balog [1] in which it is shown that for any n there are n primes $p_1, \ldots, p_n$ such that *all* of the averages $\frac{1}{2}(p_i + p_j)$ are prime. In this paper we propose to prove a common generalization of the results of Roth and Van der Corput. Write $\mathcal{P}$ for the set of primes.

THEOREM 1.4. *Every subset of $\mathcal{P}$ of positive upper density contains a* 3AP.

In fact, we get an explicit upper bound on the density of a 3AP-free subset of the primes, but it is ridiculously weak. Observe that as an immediate consequence of Theorem 1.4 we obtain what might be termed a van der Waerden theorem in the primes, at least for progressions of length 3. That is, if one colours the primes using finitely many colours then one may find a monochromatic 3AP.

We have not found a written reference for the question answered by Theorem 1.4, but M. N. Huxley has discussed it with several people [16].

To prove Theorem 1.4 we will use a variant of the following result. This says that the primes enjoy what is known as the Hardy-Littlewood majorant property.

THEOREM 1.5. *Suppose that $p \geqslant 2$ is a real number, and let $\mathcal{P}_N = \mathcal{P} \cap [1, N]$. Let $\{a_n\}_{n \in \mathcal{P}_N}$ be any sequence of complex numbers with $|a_n| \leqslant 1$ for all n. Then*

$$\left\| \sum_{n \in \mathcal{P}_N} a_n e(n\theta) \right\|_{L^p(\mathbb{T})} \leqslant C(p) \left\| \sum_{n \in \mathcal{P}_N} e(n\theta) \right\|_{L^p(\mathbb{T})}, \tag{1.1}$$

where the constant $C(p)$ depends only on p.

It is perhaps surprising to learn that such a property does not hold with *any* set $\Lambda \subseteq [N]$ in place of $\mathcal{P}_N$. Indeed, when p is an even integer it is

[1] In April 2004 the author and T. Tao published a preprint showing that the primes contain arbitrarily long arithmetic progressions.

rather straightforward to check that any set *does* satisfy (1.1) (with $C(p) = 1$). However, there are sets for which (1.1) fails badly when p is not an even integer. For a discussion of this see [10] and for related matters including connections with the Kakeya problem, see [18], [20].

We will apply a variant of Theorem 1.5 for $p = 5/2$, when it certainly does not seem to be trivial. To prove it, we will establish a somewhat stronger result which we call a *restriction theorem for primes*. The reason for this is that our argument is very closely analogous to an argument of Tomas and Stein [24] concerning Fourier transforms of measures supported on spheres.

A proof of the restriction theorem for primes was described, in a different context, by Bourgain [4]. Our argument, being visibly analogous to the approach of Tomas, is different and has more in common with Section 3 of [5]. This more recent paper of Bourgain deals with restriction phenomena of certain sets of lattice points.

To deduce Theorem 1.4 from (a variant of) Theorem 1.5 we use a variant of the technique of *granularization* as developed by I. Z. Ruzsa and the author in a series of papers beginning with [9], as well as a "statistical" version of Roth's theorem due to Varnavides. We will also require an argument of Marcinkiewicz and Zygmund which allows us to pass from the continuous setting in results such as (1.1) – that is to say, $\mathbb{T}$ – to the discrete, namely $\mathbb{Z}/N\mathbb{Z}$.

Finally, we would like to remark that it is possible, indeed probable, that Roth's theorem in the primes is true on grounds of density alone. The best known lower bound on $r_3(N)$ comes from a result of Behrend [3] from 1946.

Proposition 1.6 (Behrend). *$r_3(N) \geqslant Ne^{-C\sqrt{\log N}}$ for some absolute constant C.*

This may well give the correct order of magnitude for $r_3(N)$, and if anything like this could be proved Theorem 1.4 would of course follow trivially.

2. Preliminaries and an outline of the argument

Although the main results of this paper concern the primes in $[N]$, it turns out to be necessary to consider slightly more general sets. Let $m \leqslant \log N$ be a positive integer and let b, $0 \leqslant b \leqslant m-1$, be coprime to m. We may then define a set

$$\Lambda_{b,m,N} = \{n \leqslant N \mid nm + b \text{ is prime}\}.$$

We expect $\Lambda_{b,m,N}$ to have size about $mN/\phi(m)\log N$, and so it is natural to define a function $\lambda_{b,m,N}$ supported on $\Lambda_{b,m,N}$ by setting

$$\lambda_{b,m,N}(n) = \begin{cases} \phi(m)\log(nm+b)/mN & \text{if } n \in \Lambda_{b,m,N} \\ 0 & \text{otherwise.} \end{cases}$$

For simplicity we write $X = \Lambda_{b,m,N}$ for the next few pages. We will abuse notation and consider $\lambda_{b,m,N}$ as a measure on X. Thus for example $\lambda_{b,m,N}(X)$, which is defined to be $\sum_n \lambda_{b,m,N}(n)$, is roughly 1 by the prime number theorem in arithmetic progressions. We use $L^p(d\lambda_{b,m,N})$ norms and also the inner product $\langle f, g\rangle_X = \sum f(n)\overline{g(n)}\lambda_{b,m,N}(n)$ without further comment.

It is convenient to use the wedge symbol for the Fourier transforms on both $\mathbb{T}$ and $\mathbb{Z}$, which we define by $f^\wedge(n) = \int f(\theta)e(-n\theta)\,d\theta$ and $g^\wedge(\theta) = \sum_n g(n)e(n\theta)$ respectively. Here, of course, $e(\alpha) = e^{2\pi i\alpha}$.

For any measure space Y let $B(Y)$ denote the space of continuous functions on Y and define a map $T : B(X) \to B(\mathbb{T})$ via

$$T : f \longmapsto (f\lambda_{b,m,N})^\wedge. \tag{2.1}$$

The object of this section is to give a new proof of the following result, which may be called a *restriction theorem for primes*.

Theorem 2.1 (Bourgain). *Suppose that $p > 2$ is a real number. Then there is a constant $C(p)$ such that for all functions $f : X \to \mathbb{C}$,*

$$\|Tf\|_p \leqslant C(p)N^{-1/p}\|f\|_2. \tag{2.2}$$

Remember that the L^2 norm is taken with respect to the measure $\lambda_{b,m,N}$. Theorem 2.1 probably has most appeal when $b = m = 1$, in which case we may derive consequences for the primes themselves. Later on, however, we will take m to be a product of small primes, and so it is necessary to have the more general form of the theorem.

We turn now to an outline of the proof of Theorem 2.1. The analogy between our proof and an argument by Tomas [24], giving results of a similar nature for spheres in high-dimensional Euclidean spaces, is rather striking. In fact, the reader may care to look at the presentation of Tomas's proof in [23], whereupon she will see that there is an almost exact correspondence between the two arguments.

To begin with, the proof proceeds by the method of T and T^*, a basic technique in functional analysis. One can check that the operator $T^* : B(\mathbb{T}) \to B(X)$ is given by

$$T^* : g \longmapsto g^\wedge|_X, \tag{2.3}$$

by verifying the relation

$$\langle Tf, g\rangle_{\mathbb{T}} = \int (f\lambda_{b,m,N})^\wedge(\theta)\overline{g(\theta)}\,d\theta = \sum_n f(n)\overline{g^\wedge(n)}\lambda_{b,m,N}(n) = \langle f, T^*g\rangle_X.$$

The equation (2.3) explains the term *restriction*. Using (2.3) we see that the operator TT^* is the map from $B(\mathbb{T})$ to itself given by

$$TT^* : f \longmapsto f * \lambda^\wedge_{b,m,N}. \tag{2.4}$$

Now Theorem 2.1 may be written, in obvious notation, as

$$\|T\|_{2\to p} \leqslant C(p)N^{-1/p}. \tag{2.5}$$

The principle of T and T^*, as we will use it, states that

$$\|T\|_{2\to p}^2 = \|TT^*\|_{p'\to p} = \|T^*\|_{p'\to 2}^2. \tag{2.6}$$

We would like to emphasise that there is nothing mysterious going on here – this result is just an elegant and convenient way of bundling together some applications of Hölder's inequality. The proof of the part that we will need, that is to say the inequality $\|T\|_{2\to p}^2 \leqslant \|TT^*\|_{p'\to p}$, is simply

$$\begin{aligned}\|Tf\|_p &= \sup_{\|g\|_{p'}=1} \langle Tf, g\rangle \\ &= \sup_{\|g\|_{p'}=1} \langle f, T^*g\rangle \\ &\leqslant \|f\|_2 \sup_{\|g\|_{p'}=1} \|T^*g\|_2 \\ &= \|f\|_2 \sup_{\|g\|_{p'}=1} \langle g, TT^*g\rangle^{1/2} \\ &\leqslant \|f\|_2 \|TT^*\|_{p'\to p}^{1/2}.\end{aligned}$$

Thus we will, for much of the paper, be concerned with showing that the operator TT^* as given by (2.4) satisfies the bound

$$\|TT^*\|_{p'\to p} \leqslant C'(p)N^{-2/p}. \tag{2.7}$$

The preceding remarks show that a proof of this will imply Theorem 2.1. To get such a bound one splits λ into certain *dyadic pieces*, that is, a sum

$$\lambda_{b,m,N} = \sum_{j=1}^{K} \psi_j + \psi_{K+1}. \tag{2.8}$$

The slightly curious way of writing this indicates that the definition of ψ_{K+1} will be a little different from that of the other ψ_j. We will define these pieces so that they satisfy the L^1-L^∞ estimates

$$\|f * \psi_j^\wedge\|_\infty \ll_\varepsilon 2^{-(1-\varepsilon)j}\|f\|_1 \tag{2.9}$$

for some $\varepsilon < (p-2)/2$, and also the L^2-L^2 estimates

$$\|f * \psi_j^\wedge\|_2 \ll_\varepsilon \frac{2^{\varepsilon j}}{N}\|f\|_2. \tag{2.10}$$

Applying the Riesz-Thorin interpolation theorem (see [11, Ch. 7]) will then give

$$\|f * \psi_j^\wedge\|_p \ll 2^{-\delta j} N^{-2/p}\|f\|_{p'}$$

for some positive δ (depending on ε). Summing these estimates from $j = 1$ to $K + 1$ will establish (2.7) and hence Theorem 2.1.

To define the decomposition (2.8) we need yet more notation. From the outset we will suppose that we are trying to prove Theorem 2.1 for a particular value of p; the argument is highly and essentially nonuniform in p. Write $A = 4/(p-2)$. Let $1 < Q \leqslant (\log N)^A$. If b, m, N are as before (recall that $m \leqslant \log N$) then we define a measure $\lambda^{(Q)}_{b,m,N}$ on $\mathbb{Z}$ by setting

$$\lambda^{(Q)}_{b,m,N}(n) = \begin{cases} N^{-1} \prod_{\substack{p \leqslant Q \\ p \nmid m}} \left(1 - \frac{1}{p}\right)^{-1} & \text{if } n \leqslant N \text{ and } p \mid (nm+b) \Rightarrow p > Q \\ 0 & \text{otherwise.} \end{cases}$$

Define $\lambda^{(1)}_{b,m,N}(n) = 0$ for all n.

As Q becomes large the measures $\lambda^{(Q)}_{b,m,N}$ look more and more like $\lambda_{b,m,N}$. Much of Section 4 will be devoted to making this principle precise. We will sometimes refer to the support of $\lambda^{(Q)}_{b,m,N}$ as the set of Q-*rough* numbers.

Now let K be the smallest integer with

$$2^K > \tfrac{1}{10}(\log N)^A \tag{2.11}$$

and define

$$\psi_j = \lambda^{(2^j)}_{b,m,N} - \lambda^{(2^{j-1})}_{b,m,N} \tag{2.12}$$

for $j = 1, \ldots, K$ and define

$$\psi_{K+1} = \lambda_{b,m,N} - \lambda^{(2^K)}_{b,m,N}, \tag{2.13}$$

so that (2.8) holds. In the next two sections we prove the two required estimates, (2.9) and (2.10).

Let us note here that the main novelty in our proof of Theorem 2.1 lies in the definition of the dyadic decomposition (2.8). By contrast, the analogous dyadic decompositions in [5] take place on the Fourier side, requiring the introduction of various smooth cutoff functions not specifically related to the underlying arithmetic structure.

3. An L^2-L^2 estimate

It turns out that the proof of (2.10), the L^2-L^2 estimate, is by far the easier of the two estimates required. We have

$$\begin{aligned} \|f * \psi_j^\wedge\|_2 &= \|\widehat{f}\psi_j\|_2 \\ &\leqslant \|\psi_j\|_\infty \|\widehat{f}\|_2 \\ &= \|\psi_j\|_\infty \|f\|_2. \end{aligned}$$

Suppose first of all that $1 \leqslant j \leqslant K$. Then

$$\begin{aligned}\|\psi_j\|_\infty &\leqslant \|\lambda_{b,m,N}^{(2^j)}\|_\infty + \|\lambda_{b,m,N}^{(2^{j-1})}\|_\infty \\ &= N^{-1} \prod_{\substack{p \leqslant 2^{j+1} \\ p \nmid m}} \left(1 - \frac{1}{p}\right)^{-1} + N^{-1} \prod_{\substack{p \leqslant 2^{j} \\ p \nmid m}} \left(1 - \frac{1}{p}\right)^{-1}.\end{aligned}$$

The two products here may be estimated using Merten's formula [14, Ch. 22]:

$$\prod_{p \leqslant Q} (1 - p^{-1}) \sim \frac{e^{-\gamma}}{\log Q}.$$

This gives

$$\|\psi_j\|_\infty \ll j/N, \tag{3.1}$$

and hence

$$\|f * \psi_j^\wedge\|_2 \ll \frac{j}{N} \|f\|_2, \tag{3.2}$$

which is certainly of the requisite form (2.10). For $j = K + 1$ we have

$$\begin{aligned}\|\psi_{K+1}\|_\infty &\leqslant \|\lambda_{b,m,N}^{(2^K)}\|_\infty + \|\lambda_{b,m,N}\|_\infty \\ &\ll \log N/N,\end{aligned}$$

so that

$$\|f * \psi_{K+1}^\wedge\|_2 \ll \frac{\log N}{N} \|f\|_2. \tag{3.3}$$

This also constitutes an estimate of the type (2.10) for some $\varepsilon < (p-2)/2$. Indeed, recalling our choice of A and K (viz. (2.11)) one can check that $2^K \geqslant (\log N)^{1/\varepsilon}$ for some such ε.

4. An L^1-L^∞ estimate

This section is devoted to the rather lengthy task of proving estimates of the form (2.9).

Introduction. The first step towards obtaining an estimate of the form (2.9) is to observe that

$$\|f * \psi_j^\wedge\|_\infty \leqslant \|\psi_j^\wedge\|_\infty \|f\|_1. \tag{4.1}$$

We will prove that $\|\psi_j^\wedge\|_\infty$ is not too large by proving

PROPOSITION 4.1. *Suppose that $Q \leqslant (\log N)^A$. Then we have the estimate*

$$\|\lambda_{b,m,N}^\wedge - \lambda_{b,m,N}^{(Q)\wedge}\|_\infty \ll \log\log Q/Q.$$

The detailed proof of this fact will occupy us for several pages. Let us begin, however, by using (4.1) to see how it implies an estimate of the form (2.9). If $1 \leqslant j \leqslant K$ then,

$$\begin{aligned} \|\psi_j^\wedge\|_\infty &= \|\lambda_{b,m,N}^{(2^j)\wedge} - \lambda_{b,m,N}^{(2^{j-1})\wedge}\|_\infty \\ &\leqslant \|\lambda_{b,m,N}^\wedge - \lambda_{b,m,N}^{(2^j)\wedge}\|_\infty + \|\lambda_{b,m,N}^\wedge - \lambda_{b,m,N}^{(2^{j-1})\wedge}\|_\infty \\ &\leqslant \log j/2^j. \end{aligned} \tag{4.2}$$

This is certainly of the form (2.9). The estimate for $j = K+1$ is even easier, being immediate from Proposition 4.1.

To prove Proposition 4.1 we will use the Hardy-Littlewood circle method. Thus we divide $\mathbb{T}$ into two sets, traditionally referred to as the major and minor arcs. It is perhaps best if we define these explicitly at the outset. Thus let p be the exponent for which we are trying to prove Theorem 2.1. Recall that $A = 4/(p-2)$, and set $B = 2A + 20$. These numbers will be fixed throughout the proof. By Dirichlet's theorem on approximation, every $\theta \in \mathbb{T}$ satisfies

$$\left|\theta - \frac{a}{q}\right| \leqslant \frac{(\log N)^B}{qN} \tag{4.3}$$

for some $q \leqslant N(\log N)^{-B}$ and some a, $(a,q) = 1$. The major arcs consist of those θ for which q can be taken to be at most $(\log N)^B$. We will write this collection using the notation

$$\mathfrak{M} = \bigcup_{\substack{q \leqslant (\log N)^B \\ (a,q)=1}} \mathfrak{M}_{a,q}.$$

For these θ, the Fourier transforms $\lambda_{b,m,N}^{(Q)\wedge}$ and $\lambda_{b,m,N}^\wedge$ depend on the distribution of the almost-primes and primes along arithmetic progressions with common difference at most $(\log N)^B$. The minor arcs $\mathfrak{m}$ consist of all other θ. Here different techniques apply, and one can conclude that both $\lambda_{b,m,N}^{(Q)\wedge}$ and $\lambda_{b,m,N}^\wedge$ are small. The triangle inequality then applies.

The ingredients are as follows. The almost-primes are eminently suited to applications of sieve techniques. To keep the paper as self-contained as possible, we will follow Gowers [8] and use the arguably simplest sieve, that due to Brun, on both the major and minor arcs.

The genuine primes, on the other hand, are harder to deal with. Here we will quote two well-known results from the literature. The information concerning distribution along arithmetic progressions to small moduli comes from the prime number theorem of Siegel and Walfisz.

PROPOSITION 4.2 (Siegel-Walfisz). *Suppose that* $q \leqslant (\log N)^B$, *that* $(a,q)=1$ *and that* $1 \leqslant N_1 \leqslant N_2 \leqslant N$. *Then*

$$\sum_{\substack{N_1<p\leqslant N_2\\ p\equiv a(\mathrm{mod}\, q)}} \log p = \frac{N_2-N_1}{\phi(q)} + O\left(N\exp(-C_B\sqrt{\log N})\right). \tag{4.4}$$

The rather strange formulation of the theorem reflects the fact that the constant C_B is *ineffective* for any $B \geqslant 1$ due to the possible existence of a *Siegel zero*. For more information, including a complete proof of Proposition 4.2, see Davenport's book [7].

The techniques for dealing with the minor arcs are associated with the names of Weyl, Vinogradov and Vaughan.

The major arcs. We will have various functions $f : [N] \to \mathbb{R}$ with

$$\|f\|_\infty = O(\log N/N) \tag{4.5}$$

which are regularly distributed along arithmetic progressions in the following sense. If $L \geqslant N(\log N)^{-2B-A-1}$ and if $X \subseteq [N]$ is an arithmetic progression $\{r, r+q, \ldots, r+(L-1)q\}$ with $q \leqslant (\log N)^B$ then

$$\sum_{n\in X} f(n) = \frac{L}{N}\left(\gamma_{r,q}(f) + O((\log N)^{-A})\right), \tag{4.6}$$

where $\gamma_{r,q}$ depends only on r and q, $|\gamma_{r,q}| \leqslant q$ and the implied constant in the O term is absolute. This information is enough to get asymptotics for $f^\wedge(\theta)$ when $|\theta - a/q|$ is small, as we prove in the next few lemmas.

For a residue r modulo q, write N_r for the set $\{n \leqslant N : n \equiv r(\mathrm{mod}\, q)\}$. Write τ for the function on $\mathbb{T}$ defined by $\tau(\theta) = N^{-1}\sum_{n\leqslant N} e(\theta n)$. The first lemma deals with $f^\wedge(\theta)$ for $|\theta| \leqslant (\log N)^B/qN$.

LEMMA 4.3. *Let* r *be a residue modulo* q, *suppose that* $|\theta| \leqslant (\log N)^B/qN$, *and suppose that the function* f *satisfies* (4.5) *and* (4.6). *Then*

$$\sum_{n\in N_r} f(n)e(\theta n) = q^{-1}\gamma_{r,q}(f)\tau(\theta) + O(q^{-1}(\log N)^{-A}).$$

Proof. Set $L = N(\log N)^{-2B-A-1}$ and partition N_r into arithmetic progressions $(X_i)_{i=1}^T$ of common difference q and length between L and $2L$, where

$T \leqslant 2N/Lq$. For each i fix an element $x_i \in X_i$.

(4.7)

$$\begin{aligned}\sum_{n\in N_r} f(n)e(\theta n) &= \sum_{i=1}^{T}\sum_{n\in X_i} f(n)e(\theta n)\\ &= \sum_{i=1}^{T} e(\theta x_i)\sum_{n\in X_i} f(n) + \sum_{i=1}^{T}\sum_{n\in X_i} f(n)\left(e(\theta n)-e(\theta x_i)\right)\\ &= \sum_{i=1}^{T} e(\theta x_i)\frac{|X_i|}{N}\left(\gamma_{r,q}(f)+O((\log N)^{-A})\right)\\ &\quad +O(LN^{-1}q^{-1}(\log N)^{B+1})\\ &= \gamma_{r,q}(f)\sum_{i=1}^{T} e(\theta x_i)\frac{|X_i|}{N} + O\left(q^{-1}(\log N)^{-A}\right).\end{aligned}$$

However

$$\begin{aligned}\sum_{i=1}^{T} e(\theta x_i)|X_i| &= \sum_{i=1}^{T}\sum_{n\in X_i} e(\theta n) + \sum_{i=1}^{T}\sum_{n\in X_i}\left(e(\theta x_i)-e(\theta n)\right)\\ &= \sum_{n\in N_r} e(n\theta) + O(Lq^{-1}(\log N)^{B}).\end{aligned} \tag{4.8}$$

Finally, observe that if $0 \leqslant r, s \leqslant q-1$ then

$$\sum_{n\in N_r} e(\theta n) - \sum_{n\in N_s} e(\theta n) = O((\log N)^{B}),$$

and so

$$\left| N^{-1}\sum_{n\in N_r} e(\theta n) - q^{-1}\tau(\theta)\right| = O(N^{-1}(\log N)^{B}).$$

Combining this with (4.7) and (4.9) completes the proof of the lemma. □

We may now get an asymptotic for $f^{\wedge}(\theta)$ when θ is in the neighbourhood of a/q.

Lemma 4.4. *Suppose that f satisfies the conditions* (4.5) *and* (4.6) *and that $\theta \in \mathfrak{M}_{a,q}$ for some a, q with $(a,q)=1$ and $q \leqslant (\log N)^{B}$. Write*

$$\sigma_{a,q}(f) = \sum_{r} e(ar/q)\gamma_{r,q}(f). \tag{4.9}$$

Then,

$$f^{\wedge}(\theta) = q^{-1}\sigma_{a,q}(f)\tau(\theta - a/q) + O((\log N)^{-A}). \tag{4.10}$$

Proof. Write $\beta = \theta - a/q$. Then

$$\begin{aligned} f^\wedge(\theta) &= \sum_{n \leqslant N} f(n) e(\theta n) \\ &= \sum_{r (\mathrm{mod}\, q)} e(ar/q) \sum_{n \in N_r} f(n) e(\beta n) \\ &= q^{-1} \tau(\beta) \sum_{r (\mathrm{mod}\, q)} e(ar/q) \gamma_{r,q}(f) + O((\log N)^{-A}) \\ &= q^{-1} \sigma_{a,q}(f) \tau(\beta) + O((\log N)^{-A}). \end{aligned}$$

This concludes the proof of the lemma. □

To apply these lemmas, we need to show that $f = \lambda^{(Q)}_{b,m,N}$ and $f = \lambda_{b,m,N}$ satisfy (4.5) and (4.6) for suitable choices of $\gamma_{r,q}(f)$. We will then evaluate the sums $\sigma_{a,q}(f)$. This slightly tedious business is the subject of our next four lemmas.

LEMMA 4.5. *$f = \lambda_{b,m,N}$ satisfies* (4.5) *and* (4.6) *with*

$$\gamma_{r,q}(f) = \begin{cases} \phi(m)q/\phi(mq) & \text{if } (mr+b, mq) = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Proof. This is a fairly immediate consequence of the Siegel-Walfisz theorem (Proposition 4.2). Let $X = \{r, r+q, \ldots, r+(L-1)q\}$ be any progression contained in $[N]$ with common difference $q \leqslant (\log N)^B$ and length $L \geqslant N(\log N)^{-2B-A-1}$. An element $r + jq \in X$ lies in $\Lambda_{b,m,N}$ precisely if $(mr+b)+jmq$ is prime, so the lemma is trivially true unless $(mr+b, mq) = 1$. Supposing this to be the case, we may use Proposition 4.2. Recalling that $m \leqslant \log N$, one has

$$\begin{aligned} \lambda_{b,m,N}(X) &= \frac{\phi(m)qL}{\phi(mq)N} + O\left(mq \exp(-C_{B+1}\sqrt{\log mqN})\right) \\ &= \frac{L}{N}\left(\frac{\phi(m)q}{\phi(mq)} + O((\log N)^{-A})\right), \end{aligned}$$

as required. □

LEMMA 4.6. *$f = \lambda^{(Q)}_{b,m,N}$ satisfies* (4.5) *and* (4.6) *with*

$$\gamma_{r,q}(f) = \begin{cases} \prod_{\substack{p \leqslant Q \\ p \nmid m}} \left(1 - \frac{1}{p}\right)^{-1} \prod_{\substack{p \leqslant Q \\ p \nmid mq}} \left(1 - \frac{1}{p}\right) & \text{if } (mr+b, mq) \text{ is } Q\text{-rough} \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Consider an arithmetic progression $X = \{r, r+q, \ldots, r+(L-1)q\}$. Let $p_1, \ldots, p_k$ be the primes with $p \leqslant Q$ and $p \nmid m$. If $(mr+b, mq)$ is not Q-rough then $p_i | (mr+b, mq)$ for some i, and the second alternative of the lemma clearly holds. Suppose then that $(mr+b, mq)$ is Q-rough. We will apply the Brun sieve to estimate $\lambda^{(Q)}_{b,m,N}(X)$.

Let $x \in X$ be chosen uniformly at random, and for each i let X_i be the event $p_i | (mx+b)$. Since $p_i \nmid (mr+b, mq)$, the probability of X_i is $\varepsilon_i/p_i + O(L^{-1})$, where $\varepsilon_i = 0$ if $p_i | q$ and $\varepsilon_i = 1$ otherwise. Now we have

$$\frac{N}{L} \prod_{\substack{p \leqslant Q \\ p \nmid m}} \left(1 - \frac{1}{p}\right) \lambda^{(Q)}_{b,m,N}(X) = \mathbb{P}\left(\bigcap X_i^c\right) = U, \tag{4.11}$$

say. By the inclusion-exclusion formula it follows that for every positive integer t

$$U = \sum_{s=0}^{t} (-1)^s \sum_{1 \leqslant i_1 < \cdots < i_s \leqslant k} \prod_{j=1}^{s} \varepsilon_{i_j}/p_{i_j} + O(L^{-1}) \sum_{s=1}^{t} \binom{k}{s}. \tag{4.12}$$

It is helpful to have the error term here in a more usable form. To this end, observe that it is certainly at most $O(k^t/L)$. We wish to replace the main term in (4.12) by $\prod_{i=1}^{k} (1 - \varepsilon_i/p_i)$, which is equal to the completed sum

$$\sum_{s=0}^{k} (-1)^s \sum_{1 \leqslant i_1 < \cdots < i_s \leqslant k} \prod_{j=1}^{s} \varepsilon_{i_j}/p_{i_j}.$$

Doing this introduces an error

$$E = \sum_{s=t+1}^{k} (-1)^s \sum_{1 \leqslant i_1 < \cdots < i_s \leqslant k} \prod_{j=1}^{s} \varepsilon_{i_j}/p_{i_j},$$

which is bounded above by

$$\sum_{s=t+1}^{k} \frac{1}{s!} \left(\sum_{i=1}^{k} \frac{1}{p_i}\right)^s. \tag{4.13}$$

By another result of Mertens one has $\sum_{i=1}^{k} p_i^{-1} \leqslant \log\log Q + O(1)$. Hence if $t \geqslant 3 \log\log Q$ then each term in (4.13) is at most one half the previous one, leading to the bound

$$|E| \leqslant \frac{2(\log\log Q)^t}{t!} \leqslant \left(\frac{4e \log\log Q}{t}\right)^t.$$

Combining all of this gives

$$U = \prod_{i=1}^{k} (1 - \varepsilon_i/p_i) + O(k^t/L) + O\left((4e \log\log Q/t)^t\right).$$

Using the trivial bound $k \leqslant Q$, and choosing $t = \log N/2A \log\log N$, one gets

$$\begin{aligned} U &= \prod_{i=1}^{k} (1 - \varepsilon_i/p_i) + O(N^{-1/4A}) \\ &= \prod_{\substack{p \leqslant Q \\ p \nmid mq}} \left(1 - \frac{1}{p}\right) + O(N^{-1/4A}). \end{aligned}$$

The lemma is immediate from this and (4.11); we have

$$\begin{aligned} \lambda^{(Q)}_{b,m,N}(X) &= \prod_{\substack{p \leqslant Q \\ p \nmid m}} \left(1 - \frac{1}{p}\right)^{-1} \cdot \frac{L}{N} \cdot \left(\prod_{\substack{p \leqslant Q \\ p \nmid mq}} \left(1 - \frac{1}{p}\right) + O(N^{-1/4A}) \right) \\ &= \frac{L}{N} \left(\gamma_{r,q} + O((\log N)^{-A}) \right), \end{aligned}$$

where $\gamma_{r,q}$ has the form claimed. □

Building on the last lemma, the next lemma gives an evaluation of $\sigma_{a,q}(\lambda^{(Q)}_{b,m,N})$ and an asymptotic for $\lambda^{(Q)\wedge}_{b,m,N}(\theta)$ when $\theta \in \mathfrak{M}_{a,q}$. If $Q \geqslant 2$ we say that a positive integer is *Q-smooth* if all of its prime divisors are at most Q. We declare there to be no 1-smooth numbers.

Lemma 4.7. *Suppose that $(a, q) = 1$. Then*

$$\sigma_{a,q}(\lambda^{(Q)}_{b,m,N}) = \begin{cases} \dfrac{q\mu(q)}{\phi(q)} e\left(-\dfrac{ab\overline{m}}{q}\right) & \text{if } (m,q) = 1 \text{ and } q \text{ is } Q\text{-smooth}; \\ 0 & \text{otherwise}, \end{cases}$$

where $\overline{m}$ is the inverse of m modulo q. If $\theta \in \mathfrak{M}_{a,q}$ then

$$\lambda^{(Q)\wedge}_{b,m,N}(\theta) = \begin{cases} \dfrac{\mu(q)}{\phi(q)} e\left(-\dfrac{ab\overline{m}}{q}\right) \tau\left(\theta - \dfrac{a}{q}\right) + O((\log N)^{-A}) & \text{if } (m,q) = 1 \text{ and} \\ & \quad q \text{ is } Q\text{-smooth}; \\ O\left((\log N)^{-A}\right) & \text{otherwise}. \end{cases}$$

Proof. Recall the definition (4.9) of $\sigma_{a,q}$, and also Lemma 4.6. We shall prove that

(4.14)
$$\sum_{\substack{r \,(\mathrm{mod}\, q) \\ (mr+b, mq) \text{ is } Q\text{-rough}}} e(ar/q) = \begin{cases} e(-ab\overline{m}/q)\mu(q) & \text{if } (m,q) = 1 \text{ and } q \text{ is } Q\text{-smooth} \\ 0 & \text{otherwise.} \end{cases}$$

Now if $p|m$ then p can never divide $mr+b$, because we are assuming that $(m,b)=1$. Let q_0 be the largest factor of q which is a product of primes p with $p \leqslant Q$ and $p \nmid m$. Then the sum (4.14) is just

$$\sum_{\substack{r (\mathrm{mod}\, q) \\ (q_0, mr+b)=1}} e(ar/q). \tag{4.15}$$

Set $q_1 = q/q_0$ and write, for each r mod q, $r = kq_0 + s$ where $0 \leqslant k \leqslant q_1 - 1$ and s is a residue mod q_0. Then the sum (4.15) is

$$\sum_{\substack{s (\mathrm{mod}\, q_0) \\ (q_0, ms+b)=1}} \sum_{k=0}^{q_1-1} e\left(\frac{a(kq_0+s)}{q}\right) = \sum_{\substack{s (\mathrm{mod}\, q_0) \\ (q_0, mr+b)=1}} e(as/q) \sum_{k=0}^{q_1-1} e(ak/q_1).$$

Now a is coprime to q and hence to q_1, and therefore the rightmost sum here vanishes unless $q_1 = 1$. This is the case precisely if $q_0 = q$, which means that $(q,m)=1$ and q is Q-smooth. In this case, the sum is

$$\sum_{\substack{s (\mathrm{mod}\, q) \\ (q, ms+b)=1}} e(as/q).$$

Set $t = ms + b$. Then this sum is just

$$\begin{aligned}\sum_{\substack{t (\mathrm{mod}\, q) \\ (q,t)=1}} e\left(\frac{a\overline{m}(t-b)}{q}\right) &= e(-ab\overline{m}/q) \sum_{(q,t)=1} e(a\overline{m}t/q) \\ &= e(-ab\overline{m}/q)\mu(q).\end{aligned}$$

This last evaluation, of what is known as a *Ramanujan Sum*, is well-known and is contained, for example, in [14]. This proves (4.14).

Now to obtain $\sigma_{a,q}$ we must simply multiply (4.14) by the factor

$$F = \prod_{\substack{p<Q \\ p \nmid m}} \left(1 - \frac{1}{p}\right)^{-1} \prod_{\substack{p \leqslant Q \\ p \nmid mq}} \left(1 - \frac{1}{p}\right)$$

appearing in Lemma 4.6. One gets zero unless $(m,q)=1$ and q is Q-smooth, in which case it is not hard to see that $F = q/\phi(q)$. This completes the evaluation of $\sigma_{a,q}(\lambda^{(Q)}_{b,m,N})$, and the claimed form for $\lambda^{(Q)\wedge}_{b,m,N}(\theta)$ is an immediate consequence of Lemma 4.4. □

We need a version of the above lemma in which $\lambda^{(Q)}_{b,m,N}$ is replaced by $\lambda_{b,m,N}$. Fortunately, we can save ourselves some work by noticing that for fixed q and m we have

$$\gamma_{r,q}(\lambda_{b,m,N}) = \gamma_{r,q}(\lambda^{(Q)}_{b,m,N}) \tag{4.16}$$

for sufficiently[2] large Q. Thus $\sigma_{a,q}(\lambda_{b,m,N})$ can be evaluated by simply letting $Q \to \infty$ in the first formula of Lemma 4.7. We get

$$\sigma_{a,q}(\lambda_{b,m,N}) = \begin{cases} q\mu(q)e(-ab\overline{m}/q) & \text{if } (q,m)=1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.17}$$

This immediately leads, via Lemma 4.4, to the following evaluation of $\lambda^{\wedge}_{b,m,N}(\theta)$.

LEMMA 4.8. *Suppose that $(a,q)=1$ and that $\theta \in \mathfrak{M}_{a,q}$. Then*

$$\lambda^{\wedge}_{b,m,N}(\theta) = \begin{cases} \dfrac{\mu(q)}{\phi(q)} e\left(-\dfrac{ab\overline{m}}{q}\right) \tau\left(\theta - \dfrac{a}{q}\right) + O\left((\log N)^{-A}\right) & \textit{if } (m,q)=1 \\ O\left((\log N)^{-A}\right) & \textit{otherwise.} \end{cases} \tag{4.18}$$

The minor arcs. In this subsection we look at $\lambda^{\wedge}_{b,m,N}(\theta)$ and $\lambda^{(Q)\wedge}_{b,m,N}(\theta)$ when θ is not close to a rational with small denominator.

LEMMA 4.9. *Suppose that a, q are positive integers with $(a,q)=1$, and let θ be a real number such that $|\theta - a/q| \leqslant 1/q^2$. Then*

$$\lambda^{\wedge}_{b,m,N}(\theta) \ll (\log N)^{10}\left(q^{-1/2} + N^{-1/5} + N^{-1/2}q^{1/2}\right). \tag{4.19}$$

Thus if $\theta \in \mathfrak{m}$ then $\lambda^{\wedge}_{b,m,N}(\theta) = O((\log N)^{-A})$.

Remarks. This is a well-known estimate, at least when $b = m = 1$. The first (unconditional) results of this type were obtained by I.M. Vinogradov, and nowadays it is possible to give a rather clean argument thanks to the identity of Vaughan [26]. Chapter 24 of Davenport's book [7] describes the use of Vaughan's identity in the more general context of the estimation of sums $\sum_{n \leqslant N} \Lambda(n) f(n)$. To obtain Lemma 4.9 we used this approach, but could afford to obtain results which are rather nonuniform in m due to the restriction $m \leqslant \log N$ under which we are operating. Details may be found in the supplementary document [12]. We remark that existing results in the literature concerning minor arcs estimates for primes restricted to arithmetic progressions, such as [2], [17], strive for a much better dependence on the parameter m.

[2]Here we regard $\gamma_{r,q}(\lambda_{b,m,N})$ and $\gamma_{r,q}(\lambda^{(Q)}_{b,m,N})$ as purely formal expressions, so there is no issue of whether or not, for example, Lemma 4.7 is valid for "sufficiently large" Q.

LEMMA 4.10. *Suppose that a, q are positive integers with $(a,q)=1$, and let θ be a real number such that $|\theta - a/q| \leqslant 1/q^2$. Then*

$$\lambda^{(Q)\wedge}_{b,m,N}(\theta) \ll (\log N)^3 \left(q^{-1} + qN^{-1} + N^{-1/8A}\right). \tag{4.20}$$

Thus if $\theta \in \mathfrak{m}$ then $\lambda^{(Q)\wedge}_{b,m,N}(\theta) = O((\log N)^{-A})$.

Proof. Let $p_1, \ldots, p_k$ be the primes less than or equal to Q which do not divide m. Another application of the inclusion-exclusion principle gives

$$\lambda^{(Q)\wedge}_{b,m,N}(\theta) = N^{-1} e(-b\theta/m) \prod_{i=1}^{k} \left(1 - \frac{1}{p_i}\right)^{-1} h(\theta),$$

where

$$h(\theta) = \sum_{s=0}^{k} (-1)^s \sum_{1 \leqslant i_1 < \cdots < i_s \leqslant k} \sum_{\substack{1 \leqslant y \leqslant Nm/p_{i_1} \ldots p_{i_s} \\ y \equiv b (\mathrm{mod}\, m)}} e\left(\frac{\theta p_{i_1} \ldots p_{i_s} y}{m}\right). \tag{4.21}$$

Summing the geometric progression, one sees that the inner sum is no more than

$$\min\left\{\|\theta p_{i_1} \ldots p_{i_s}\|^{-1}, 2mN/p_{i_1} \ldots p_{i_s}\right\}.$$

We will split the sum over s in (4.21) into two pieces, over the ranges $s \in [0,t]$ and $s \in (t,k]$ where $t = \log N/2A \log\log N$. Each of the primes p_i is at most $Q \leqslant (\log N)^A$, so the product of any $s \leqslant t$ of them is no more than $\sqrt{N}$. Of course, all such products are distinct and so

$$\sum_{s=0}^{t} (-1)^s \sum_{1 \leqslant i_1 < \cdots < i_s \leqslant k} \sum_{\substack{y \leqslant Nm/p_{i_1} \ldots p_{i_s} \\ y \equiv b (\mathrm{mod}\, m)}} e\left(\frac{\theta p_{i_1} \ldots p_{i_s} y}{m}\right) \leqslant \sum_{n \leqslant \sqrt{N}} \min(\|\theta n\|^{-1}, 2mN/n).$$

This is a quantity whose estimation is standard in this area because of its pertinence to the estimation of exponential sums on minor arcs. It is bounded above by $C(\log N)^3(N^{1/2} + q + Nq^{-1})$; details may once again be found in [12].

On the other hand

$$\begin{aligned}\sum_{s=t+1}^{k}(-1)^s \sum_{1\leqslant i_1<\cdots<i_s\leqslant k}\ \sum_{\substack{y\leqslant Nm/p_{i_1}\ldots p_{i_s}\\ y\equiv b(\operatorname{mod} m)}} e\left(\frac{\theta p_{i_1}\ldots p_{i_s}y}{m}\right)\\ \leqslant 2mN\sum_{s=t+1}^{k}\sum_{1\leqslant i_1<\cdots<i_s\leqslant k}\prod_{j=1}^{s}p_{i_j}^{-1}\\ \leqslant 2mN\sum_{s=t+1}^{k}(s!)^{-1}\left(p_1^{-1}+\cdots+p_k^{-1}\right)^s\\ \leqslant 4mN(2e\log\log\log N/t)^t\leqslant mN^{1-1/4A}\leqslant N^{1-1/8A}.\end{aligned}$$

Since $\prod_{i=1}^{k}(1-1/p_i)^{-1}\ll\log N$, the claimed bound follows. □

Proof of Proposition 4.1. Suppose first of all that $\theta\in\mathfrak{M}_{a,q}$ for some a,q, and recall Lemmas 4.7 and 4.8. If q is Q-smooth then

$$\left|\lambda^{\wedge}_{b,m,N}(\theta)-\lambda^{(Q)\wedge}_{b,m,N}(\theta)\right|=O(N(\log N)^{-A}).$$

If q is not Q-smooth then $q>Q$ and so we get

$$\begin{aligned}\left|\lambda^{\wedge}_{b,m,N}(\theta)-\lambda^{(Q)\wedge}_{b,m,N}(\theta)\right|&\leqslant|\lambda^{\wedge}_{b,m,N}(\theta)|+|\lambda^{(Q)\wedge}_{b,m,N}(\theta)|\\ &\leqslant 2/\phi(q)+O((\log N)^{-A})\\ &\leqslant 4\log\log Q/Q+O((\log N)^{-A}),\end{aligned}$$

the last estimate being contained in [14, Ch. 7]. Since we are assuming that $Q\leqslant(\log N)^A$ this expression is $O(\log\log Q/Q)$. If, on the other hand, $\theta\in\mathfrak{m}$ then we have

$$\begin{aligned}\left|\lambda^{\wedge}_{b,m,N}(\theta)-\lambda^{(Q)\wedge}_{b,m,N}(\theta)\right|&\leqslant|\lambda^{\wedge}_{b,m,N}(\theta)|+|\lambda^{(Q)\wedge}_{b,m,N}(\theta)|\\ &=O((\log N)^{-A})\\ &=O(Q^{-1}).\end{aligned}$$

This at last completes the proof of Proposition 4.1. □

5. Restriction and majorant estimates for primes

In this section we prove Theorems 1.5 and 2.1.

We have already seen, in (4.1) and (4.2), how Proposition 4.1 implies an L^1-L^∞ estimate for the operator $f\mapsto f*\psi_j$ of the form (2.9). In fact, we have

$$\|f*\psi_j\|_\infty\ll\frac{\log j}{2^j}\|f\|_1 \tag{5.1}$$

for all $j=1,\ldots,K+1$. For each fixed $j=1,\ldots,K$, one can use the Riesz-Thorin interpolation theorem to interpolate between (3.2) and (5.1). This

theorem, which is discussed in [11, Ch. 7], is better known to analytic number theorists as the type of convexity principle that underpins many basic estimates on ζ and L-functions. It gives

$$\|f * \psi_j\|_p \ll j^{2/p}(\log j)^{1-2/p}2^{-(1-2/p)j}N^{-2/p}\|f\|_{p'}. \tag{5.2}$$

For $j = K + 1$ another interpolation, now between (3.3) and (5.1), instead gives

$$\|f * \psi_{K+1}\|_p \ll (\log N)^{2/p}(\log K)^{1-2/p}2^{-(1-2/p)K}.$$

Recalling at this point the definition (2.11) of K we see that this implies

$$\|f * \psi_{K+1}\|_p \ll (\log N)^{-1/p}N^{-2/p}.$$

Summing this together with (5.2) for $j = 1, \ldots, K$ gives, because of the decomposition (2.8),

$$\|f * \lambda_{b,m,N}\|_p \leqslant C(p)N^{-2/p}\|f\|_{p'}.$$

As we have already remarked, Theorem 2.1 follows by the principle of T and T^*. □

Now we prove Theorem 1.5. Although we will need a slightly different result later on, this theorem seems to be the most elegant way to state the majorant property for the primes.

Proof of Theorem 1.5. Let $(a_n)_{n\in\mathcal{P}_N}$ be any sequence of complex numbers with $|a_n| \leqslant 1$ for all n. We apply Theorem 2.1 to the function f defined by $f(n) = a_n/\log n$. Writing out the conclusion of Theorem 2.1 gives, for any $p > 2$,

$$\int \left|\sum_n f(n)\log n e(n\theta)\right|^p d\theta \ll_p N^{p/2-1}\left(\sum_n f(n)^2\log n\right)^{p/2}.$$

Therefore

$$\int \left|\sum_{n\in\mathcal{P}_N} a_n e(n\theta)\right|^p d\theta \ll_p N^{p/2-1}\left(\sum_{n\in\mathcal{P}_N}\frac{|a_n|^2}{\log n}\right)^{p/2}$$
$$\ll_p N^{p-1}(\log N)^{-p}.$$

However it is an easy matter to check that

$$\int \left|\sum_{n\in\mathcal{P}_N} e(n\theta)\right|^p d\theta \geqslant \int_{|\theta|\leqslant 1/2N}\left|\sum_{n\in\mathcal{P}_N} e(n\theta)\right|^p d\theta \gg N^{p-1}(\log N)^{-p}.$$

This proves Theorem 1.5 for $p > 2$. For $p = 2$ it is trivial by Parseval's identity. □

6. Roth's theorem in the primes

Let A_0 be a subset of the primes with positive relative upper density. By this we mean that there is a positive constant α_0 such that, for infinitely many integers n, we have

$$|A \cap \mathcal{P}_n| \geqslant \alpha_0 n/\log n. \tag{6.1}$$

This is not a particularly convenient statement to work with, and our first lemma derives something more useful from it.

LEMMA 6.1. *Suppose that there is a set $A_0 \subseteq \mathcal{P}$ with positive relative density, but which contains no* 3AP*s. Then there are a positive real number α and infinitely many primes N for which the following is true. There are a set $A \subseteq \{1, \ldots, \lfloor N/2 \rfloor\}$, and an integer $W \in [\frac{1}{8}\log\log N, \frac{1}{4}\log\log N]$ such that*

- *A contains no* 3AP*s,*
- *$\lambda_{b,m,N}(A) \geqslant \alpha$ for some b with $(b,m) = 1$, where $m = \prod_{p \leqslant W} p$.*

Proof. Take any $n \geqslant \alpha_0^{-3}$ for which (6.1) holds. Let $W = \lfloor \frac{1}{4}\log\log n \rfloor$, and set $m = \prod_{p \leqslant W} p$. Choose N to be any prime in the range $(2n/m, 4n/m]$. Now there are certainly no more than m elements of A_0 which share a factor with m, and no more than $n^{3/4}$ elements $x \in A_0$ with $x \leqslant n^{3/4}$. Thus

$$\sum_{b:(b,m)=1} \sum_{\substack{x \leqslant n \\ x \equiv b (\mathrm{mod}\, m)}} A_0(x) \log x \geqslant \alpha_0 n/2,$$

and for some choice of b we have

$$\sum_{\substack{x \leqslant n \\ x \equiv b (\mathrm{mod}\, m)}} A_0(x) \log x \geqslant \alpha_0 n/2\phi(m). \tag{6.2}$$

Write $A = m^{-1}((A_0 \cap [n]) - b)$. This set, being a part of A_0 subjected to a linear transformation, contains no 3-term AP. It is also clear that $A \subseteq \{1, \ldots, \lfloor N/2 \rfloor\}$. Furthermore (6.2) is equivalent to

$$\sum_{\substack{x \leqslant N \\ mx+b \text{ is prime}}} A(x) \log(mx+b) \geqslant \alpha_0 n/2\phi(m),$$

which implies that $\lambda_{b,m,N}(A) \geqslant \alpha_0 n/2mN \geqslant \alpha_0/8$. The lemma follows, with $\alpha = \alpha_0/8$. □

The reason we stipulate that A be contained in $\{1, \ldots, \lfloor N/2 \rfloor\}$ is that A does not contain any 3APs when considered *as a subset of* $\mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$. This

allows us to make use of Fourier analysis on $\mathbb{Z}_N$. If $f : \mathbb{Z}_N \to \mathbb{C}$ is a function we will write, for any $r \in \mathbb{Z}_N$,

$$\widetilde{f}(r) = \sum_{x \in \mathbb{Z}_N} f(x) e(-rx/N).$$

Observe that f may also be considered as a function on $\mathbb{Z}$ via the embedding $\mathbb{Z}_N \hookrightarrow [N]$, and then $\widetilde{f}(r) = f^\wedge(r/N)$.

For notational simplicity write $\mu = \lambda_{b,m,N}$. We will consider A and μ as functions on $\mathbb{Z}_N$. Write $a = A\mu$. We will continue to abuse notation by using μ and a as measures. Thus, for example, $a(\mathbb{Z}_N) \geqslant \alpha$.

Now if A contains no (nontrivial) 3APs then

$$\begin{aligned} \sum_{x,d} a(x)a(x+d)a(x+2d) &= \sum_x a(x)^3 \\ &\leqslant \sum_x \mu(x)^3 \\ &\leqslant (\log N)^3/N^2. \end{aligned} \tag{6.3}$$

We are going to show that this forces α to be small. We will do this by constructing a new measure a_1 on $\mathbb{Z}_N$ which is *set-like*, which means that a_1 behaves a bit like N^{-1} times the characteristic function of a set of size $\sim \alpha N$. The new measure a_1 will be fairly closely related to a, and in fact we will be able to show that

$$\sum_{x,d} a_1(x)a_1(x+d)a_1(x+2d) \qquad \text{is small.} \tag{6.4}$$

This, it turns out, is impossible; an argument of Varnavides based on Roth's theorem tells us that a dense subset of $\mathbb{Z}_N$ contains *lots* of 3APs. We will adapt his argument in a trivial way to show that the same is true of set-like measures.

The arguments of this section, then, fall into two parts. First of all we must define a_1, define the notion of "set-like" and then show that a_1 is indeed set-like. The key ingredient here is Lemma 6.2, which says that $\widetilde{\mu}$ is small away from zero. Secondly, we must formulate and prove a result of the form (6.4). For this we need Theorem 2.1, the restriction theorem for primes.

The idea of constructing a_1, and the technique for constructing it, has its origins in the notions of *granularization* as used in a paper of I.Z. Ruzsa and the author [9]. In the present context things look rather different however and, in the absence of anything which might be called a "grain", we think the terminology of [9] no longer appropriate.

Let us proceed to the definition of a_1. Let $\delta \in (0,1)$ be a real number to be chosen later, and set

$$R = \{r \in \mathbb{Z}_N : |\widetilde{a}(r)| \geqslant \delta\}.$$

Let $k = |R|$, and write $R = \{r_1, \ldots, r_k\}$. Let $\varepsilon \in (0,1)$ be another real number to be chosen later, and write $B(R, \varepsilon)$ for the Bohr neighbourhood

$$\left\{x \in \mathbb{Z}_N : \left\|\frac{xr_i}{N}\right\| \leqslant \varepsilon \ \forall i \in [k]\right\}.$$

Write $B = B(R, \varepsilon)$ and set $\beta(x) = B(x)/|B|$. Define

$$a_1 = a * \beta * \beta. \tag{6.5}$$

It is easy to see that

$$a_1(\mathbb{Z}_N) \geqslant \alpha. \tag{6.6}$$

In Lemma 6.3 below we will show that $\|a_1\|_\infty \leqslant 2/N$, provided that a certain inequality between ε, k and W is satisfied. This is what we mean by the statement that a_1 is set-like.

Lemma 6.2. *Suppose that N, and hence W, is sufficiently large. Then,*

$$\sup_{r \neq 0} |\widetilde{\mu}(r)| \leqslant 2 \log\log W/W.$$

Proof. Recall that $\widetilde{\mu}(r) = \mu^\wedge(r/N)$. There are three different cases to consider.

Case 1. $r/N \in \mathfrak{M}_{0,1}$; that is to say $|r/N| \leqslant (\log N)^B/N$. Then by Lemma 4.8 we have the asymptotic

$$\widetilde{\mu}(r) = \tau(r/N) + O(\log N)^{-A}.$$

Observe, however, that $\tau(r/N) = 0$ provided that $r \neq 0$.

Case 2. $r/N \in \mathfrak{M}_{a,q}$. Then Lemma 4.8 gives

$$\widetilde{\mu}(r) = \frac{\chi_q \mu(q)}{\phi(q)} e\left(-\frac{ab\overline{m}}{q}\right) \tau\left(\frac{r}{N} - \frac{a}{q}\right) + O(\log N)^{-A},$$

where

$$\chi_q = \begin{cases} 1 & (q,m) = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Since $m = \prod_{p \leqslant W} p$, we certainly have $\chi_q = 0$ for $q \leqslant W$. Thus indeed

$$|\widetilde{\mu}(r)| \leqslant \sup_{n \geqslant W} \phi(n)^{-1} + O(\log N)^{-A} \leqslant 2 \log\log W/W.$$

Case 3. $r/N \in \mathfrak{m}$. Then Lemma 4.9 gives

$$\widetilde{\mu}(r) = \mu^\wedge(r/N) = O((\log N)^{-A}). \qquad \square$$

LEMMA 6.3. *Suppose that* $\varepsilon^k \geqslant 2\log\log W/W$. *Then the measure* a_1 *is set-like, in the sense that* $\|a_1\|_\infty \leqslant 2/N$.

Proof. Indeed

$$\begin{aligned} a_1(x) &= a * \beta * \beta(x) \\ &\leqslant \mu * \beta * \beta(x) \\ &= N^{-1}\sum_r \widetilde{\mu}(r)\widetilde{\beta}(r)^2 e(rx/N) \\ &\leqslant N^{-1}\widetilde{\mu}(0)\widetilde{\beta}(0)^2 + N^{-1}\sum_{r\neq 0}|\widetilde{\mu}(r)||\widetilde{\beta}(r)|^2 \\ &\leqslant N^{-1} + N^{-1}\sup_{r\neq 0}|\widetilde{\mu}(r)|\sum_r |\widetilde{\beta}(r)|^2 \\ &= N^{-1} + |B|^{-1}\sup_{r\neq 0}|\widetilde{\mu}(r)| \\ &\leqslant N^{-1} + \frac{2\log\log W}{W|B|}. \end{aligned}$$

Now by a well-known application of the pigeonhole principle we have $|B| \geqslant \varepsilon^k N$, from which the lemma follows immediately. □

We move on now to the second part of our programme, which includes a statement and proof of a result of the form (6.4).

PROPOSITION 6.4. *There is an inequality*

$$\sum_{x,d} a_1(x)a_1(x+d)a_1(x+2d) \leqslant C'N^{-3/2} + \frac{1}{N}\left(2^{12}\varepsilon^2\delta^{-5/2} + C\delta^{1/2}\right).$$

We will require several lemmas. The most important is a "discrete majorant property". Before we state and prove this, we give an elegant argument of Marcinkiewicz and Zygmund [27]. We outline the argument here since we like it and, possibly, it is not particularly well-known.

LEMMA 6.5 (Marcinkiewicz-Zygmund). *Let* N *be a positive integer, and let* $f : [N] \to \mathbb{C}$ *be any function. Consider* f *also as a function on* $\mathbb{Z}_N$. *Let* $p > 1$ *be a real number. Then*

$$\sum_{r\in\mathbb{Z}_N} |\widetilde{f}(r)|^p = \sum_{r=0}^{N-1} |f^\wedge(r/N)|^p \leqslant C(p)N\int |\widetilde{f}(\theta)|^p\, d\theta.$$

Proof. Consider the function

$$g(n) = 2\left(1 - \frac{|n|}{2N}\right)\chi_{|n|\leqslant 2N} - \left(1 - \frac{|n|}{N}\right)\chi_{|n|\leqslant N}.$$

This function is equal to 1 for all n with $|n| \leqslant N$. Its Fourier transform, $g^\wedge(\theta)$, is equal to $2K_{2N}(\theta) - K_N(\theta)$, a difference of two Fejér kernels. Thus we have

$$f^\wedge = f^\wedge * (2K_{2N} - K_N)\,,$$

and so

$$\begin{aligned}
|\widetilde{f}(r)|^p &= |f^\wedge(r/N)|^p \\
&= \left| \int f^\wedge(\theta)\,(2K_{2N}(r/N-\theta) - K_N(r/N-\theta))\,d\theta \right|^p \\
&\leqslant 3^{p-1}\left(2^p \left| \int f^\wedge(\theta) K_{2N}(r/N-\theta)\,d\theta \right|^p + \left| \int f^\wedge(\theta) K_N(r/N-\theta)\,d\theta \right|^p \right) \\
&\leqslant 3^{p-1}\left(2^p \int |f^\wedge(\theta)|^p K_{2N}(r/N-\theta)\,d\theta + \int |f^\wedge(\theta)|^p K_N(r/N-\theta)\,d\theta \right)
\end{aligned}$$

by two applications of Jensen's inequality. It is necessary, of course, to use the fact that the Fejér kernels are nonnegative. To conclude the proof, one only has to show that

$$\sum_{r=0}^{N-1} K_N(r/N-\theta) \leqslant CN,$$

together with a similar inequality for K_{2N}. But this is a straightforward matter using the bound

$$\sum_{r=0}^{N-1} K_N(r/N-\theta) \leqslant \sum_{j=0}^{N-1} \sup_{\phi \in [\frac{j}{N}, \frac{j+1}{N}]} K_N(\phi)$$

together with the estimate

$$K_N(\phi) \ll \min(N, N^{-1}|\phi|^{-2}),$$

valid for $|\phi| \leqslant 1/2$. □

Lemma 6.6 (Discrete majorant property). *Suppose that $p > 2$. Then there is an absolute constant $C(p)$ (not depending on a) such that*

$$\sum_r |\widetilde{a}(r)|^p \leqslant C(p).$$

Proof. A direct application of Theorem 2.1 gives

$$\int |a^\wedge(\theta)|^p\,d\theta \leqslant C'(p)N^{-1}.$$

The lemma is immediate from this and Lemma 6.5. □

LEMMA 6.7. *Suppose that $r \in R$. Then* $\left|1 - \widetilde{\beta}(r)^4 \widetilde{\beta}(-2r)^2\right| \leqslant 2^{12}\varepsilon^2$.

Proof. We have

$$\begin{aligned}\left|1 - \widetilde{\beta}(r)\right| &= \frac{1}{|B|}\left|\sum_{x\in B}(1 - e(rx/N))\right| \\ &= \frac{1}{|B|}\left|\sum_{x\in B}(1 - \cos(2\pi rx/N))\right| \\ &\leqslant 4\pi^2 \sup_{x\in B}\|rx/N\|^2 \\ &\leqslant 16\varepsilon^2.\end{aligned}$$

A very similar calculation shows that

$$\left|1 - \widetilde{\beta}(-2r)\right| \leqslant 64\varepsilon^2,$$

and the lemma follows quickly. □

Proof of Proposition 6.4. By (6.3) we have, observing that $\widetilde{a_1} = \widetilde{a}\widetilde{\beta}^2$,

(6.7)
$$\begin{aligned}\sum a_1(x)a_1(x+d)a_1(x+2d) &\leqslant \sum a_1(x)a_1(x+d)a_1(x+2d) \\ &\quad - \sum a(x)a(x+d)a(x+2d) + (\log N)^3 N^{-2} \\ &= O(N^{-3/2}) \\ &\quad - N^{-1}\sum_r \widetilde{a}(r)^2\widetilde{a}(-2r)\left(1 - \widetilde{\beta}(r)^4\widetilde{\beta}(-2r)^2\right).\end{aligned}$$

Split the sum in (6.7) into two parts, one over $r \in R$ and the other over $r \notin R$. When $r \in R$ we use Lemma 6.7 to get

$$\begin{aligned}\sum_{r\in R}\widetilde{a}(r)^2\widetilde{a}(-2r)\left(1 - \widetilde{\beta}(r)^4\widetilde{\beta}(-2r)^2\right) &\leqslant 2^{12}\varepsilon^2|R| \\ &\leqslant C\varepsilon^2\delta^{-5/2},\end{aligned}$$

this last inequality following from Lemma 6.6 with $p = 5/2$. To estimate the sum over $r \notin R$, we again use Lemma 6.6 with $p = 5/2$. Indeed using Hölder's inequality we have

$$\begin{aligned}\left|\sum_{r\notin R}\widetilde{a}(r)^2\widetilde{a}(-2r)\left(1 - \widetilde{\beta}(r)^4\widetilde{\beta}(-2r)^2\right)\right| &\leqslant 2\sup_{r\notin R}|\widetilde{a}(r)|^{1/2}\sum_r |\widetilde{a}(r)|^{5/2} \\ &\leqslant C\delta^{1/2}.\end{aligned}$$

This concludes the proof of Proposition 6.4. □

By (6.6) and Lemma 6.3, a_1 behaves a bit like a measure associated to a set of size αN. As promised, we use this information together with an argument originally due to Varnavides [25] to get a *lower* bound on $\sum a_1(x)a_1(x+d)a_1(x+2d)$.

LEMMA 6.8. *For some absolute constant* C_2,

$$\sum_{x,d\in\mathbb{Z}_N} a_1(x)a_1(x+d)a_1(x+2d) \geqslant \exp\left(-C_2\alpha^{-2}\log(1/\alpha)\right)N^{-1}.$$

Proof. Let $A' = \{x \in \mathbb{Z}_N : a_1(x) \geqslant \alpha/2N\}$. By Lemma 6.3 we have

$$\alpha \leqslant \sum a_1(x) \leqslant \frac{2|A'|}{N} + \frac{\alpha}{2N}|A'^c|,$$

which implies that $|A'| \geqslant \alpha N/4$. We will give a lower bound for Z, the number of 3APs in A'. It is clear that $\sum a_1(x)a_1(x+d)a_1(x+2d)$ is at least $\alpha^3 Z/8N^3$.

Now by Bourgain's theorem[3] [6] there is a constant C_1 such that if

$$M \geqslant \exp\left(C_1\alpha^{-2}\log(1/\alpha)\right)$$

then any subset of $\{1,\dots,M\}$ of density at least $\alpha/8$ contains a 3AP with nonzero common difference. Now there are exactly $N(N-1)$ nontrivial arithmetic progressions of length M in $\mathbb{Z}_N$, and A' will have density at least $\alpha/8$ on many of them. To estimate exactly how many, fix a common difference $d \neq 0$, and let $I = \{0, d, 2d, \dots, (M-1)d\}$. We have $\sum_x A' * I(x) \geqslant \alpha NM/4$, but $A' * I(x) \leqslant M$ for every x. Thus another simple averaging argument shows that $A' * I(x) \geqslant \alpha M/8$ for at least $\alpha N/8$ values of x.

In total, then, there are at least $\alpha N^2/8$ progressions of length M on which A' has density at least $\alpha/8$. Each of them contains a 3AP consisting of elements of A'. No 3AP thus counted can arise from more than M^2 progressions of length M. Thus we have two different ways of bounding Z, and putting them together gives

$$Z \geqslant \alpha N^2/8M^2.$$

The lemma follows. □

Combining this with Proposition 6.4, we get

$$C'N^{-1/2} + 2^{12}\varepsilon^2\delta^{-5/2} + C\delta^{1/2} \geqslant \exp\left(-C_2\alpha^{-2}\log(1/\alpha)\right). \tag{6.8}$$

There are constants C_3, C_4 so that if we choose

$$\delta = \exp\left(-C_3\alpha^{-2}\log(1/\alpha)\right)$$

[3] We could equally well use Roth's original theorem here, at the expense of making any bounds for the relative density in Theorem 1.4 even worse.

and

$$\varepsilon = \exp\left(-C_4\alpha^{-2}\log(1/\alpha)\right)$$

then (6.8) cannot hold, and we will have derived a contradiction to the assumption that A contains no 3APs. We are permitted to choose any values of ε and δ so that the condition of Lemma 6.8 is satisfied. Recalling that $k \leqslant C\delta^{-5/2}$ (a consequence of Lemma 6.6) and that $W \geqslant \log\log N/8$, we see that (6.8) can indeed be contradicted provided that

$$\alpha \geqslant C\sqrt{\frac{\log_5 N}{\log_4 N}}. \tag{6.9}$$

The subscripts indicate the number of iterated logarithms, not the base to which those logarithms are taken!

Let us remind the reader of what it is that we have contradicted. We assumed that there was a subset $A_0 \subseteq \mathcal{P}$ of positive relative upper density, containing no 3AP. The number α was related to the relative upper density of A_0, via the slightly technical reductions made in Lemma 6.1. A bound of the form (6.9) also holds for α_0. That is, any subset of $\mathcal{P}_n$ with cardinality at least $Cn(\log_5 n)^{1/2}/\log n(\log_4 n)^{1/2}$ contains a 3AP.

By far the most important reason for our getting such a poor bound was the need to prove Lemma 6.2, which says that by passing to a subprogression of common difference $m = \prod_{p\leqslant W} p$ one can make the primes look somewhat uniform. This is a rather crude trick but we have not been able to get around it. Even if we could, the resultant bounds would surely be many miles from the probable truth, which is that *any* subset of $[N]$ of cardinality $N(\log N)^{-1000}$ contains 3APs.

Let us conclude by remarking that the methods of this section use rather little about the primes. In fact by the same argument one could establish a Roth-type theorem relative to any measure $\mu : \mathbb{Z}_N \to \mathbb{R}^+$ for which one had good control on $\sup_{r\neq 0}|\widetilde{\mu}(r)|$ together with bounds for $\|\widetilde{f}\|_p$, for some $p \in (2,3)$ and any f satisfying $0 \leqslant f(x) \leqslant \mu(x)$ pointwise. In practise bounds of this latter type will come by restriction theory arguments of the type given in Section 5. A more general setting for our arguments, along the lines just described, is given in [13].

7. Acknowledgements

The author would like to thank Tim Gowers for his insights into Vinogradov's three-primes theorem, which played a substantial part in the development of this paper. He would also like to thank Imre Ruzsa for helpful conversations, Jean Bourgain for drawing his attention to the references [4], [5] and the students who attended the course [11] for their enthusiasm.

Trinity College, University of Cambridge, Cambridge, United Kingdom
E-mail address: bjg23@hermes.cam.ac.uk

References

[1] A. Balog, Linear equations in primes, *Mathematika* **39** (1992), 367–378.

[2] A. Balog and A. Perelli, Exponential sums over primes in an arithmetic progression, *Proc. Amer. Math. Soc.* **93** (1985), 578–582.

[3] F. A. Behrend, On sets of integers which contain no three terms in arithmetical progression, *Proc. Nat. Acad. Sci. U.S.A.* **32** (1946), 331–332.

[4] J. Bourgain, On $\Lambda(p)$-subsets of squares, *Israel J. Math.* **67** (1989), 291–311.

[5] ———, Fourier transform restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations. I. Schrödinger equations, *Geom. Funct. Anal.* **3** (1993), 107–156.

[6] ———, On triples in arithmetic progression, *Geom. Funct. Anal.* **9** (1999), 968–984.

[7] H. Davenport, *Multiplicative Number Theory*, Third edition, *Grad. Texts Math.* **74** Springer-Verlag, New York, 2000.

[8] W. T. Gowers,Vinogradov's three-primes theorem; notes available at `http://www.dpmms.cam.ac.uk/~wtg10/3primes.dvi`.

[9] B. J. Green and I. Z. Ruzsa, Counting sumsets and sum-free sets modulo a prime, *Studia Sci. Math. Hungar.* **41** (2004), 285–293.

[10] ———, On the Hardy-Littlewood majorant problem, *Math. Proc. Camb. Phil. Soci.* **137** (2004), 511–517.

[11] B. J. Green, Restriction and Kakeya phenomena, notes from a course given in Part III of the Cambridge Mathematical Tripos (2002); Available at `http://www.dpmms.cam.ac.uk/~bjg23/rkp.html`.

[12] ———, Some minor arcs estimates relevant to the paper "Roth's theorem in the primes"; preprint, available at `http://www.dpmms.cam.ac.uk/~bjg23/papers/BG_11_minorarcs.pdf`

[13] B. J. Green and T. C. Tao, Restriction theory of the Selberg sieve, with applications, preprint, available at `http://www.arxiv.org/abs/math.NT/0405581`

[14] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Fifth edition, The Clarendon Press, Oxford University Press, New York, 1979.

[15] D. R. Heath-Brown, Integer sets containing no arithmetic progressions, *J. London Math. Soc.* **35** (1987), 385–394.

[16] M. N. Huxley, Personal communication.

[17] A. F. Lavrik, Analytic method of estimates of trigonometric sums by the primes of an arithmetic progression (Russian), *Dokl. Akad. Nauk SSSR* **248** (1979), 1059–1063.

[18] G. Mockenhaupt, Bounds in Lebesgue spaces of oscillatory integrals, Habilitation thesis, Universität Siegen, 1996; Available at `http://http://www.math.gatech.edu/~gerdm/research/`.

[19] F. Mockenhaupt and T. Tao, Restriction and Kakeya phenomena in finite fields, *Duke Math. J.* **121** (2004), 35–74.

[20] G. Mockenhaupt and W. Schlag, The Hardy-Littlewood majorant property for random sets, preprint.

[21] K. F. Roth, On certain sets of integers, *J. London Math. Soc.* **28** (1953), 104–109.

[22] E. Szemerédi, Integer sets containing no arithmetic progressions, *Acta Math. Hungar.* **56** (1990), 155–158.

[23] T. Tao, Notes from a course given at UCLA, available at `http://www.math.ucla.edu/~tao/254b.1.99s/notes1.dvi`.

[24] P. Tomas, A restriction theorem for the Fourier transform, *Bull. Amer. Math. Soc.* **81** (1975), 477–478.

[25] P. Varnavides, On certain sets of positive density, *J. London Math. Soc.* **34** (1959), 358–360.

[26] R. C. Vaughan, Sommes trigonométriques sur les nombres premiers, *C. R. Acad. Sci. Paris Sér. A-B* **285** (1977), A981–A983.

[27] A. Zygmund, *Trigonometric Series*, 2nd ed., Vols. I, II, Cambridge Univ. Press, New York, 1959.

(Received February 27, 2003)