LARGE GAPS BETWEEN CONSECUTIVE PRIME NUMBERS

KEVIN FORD, BEN GREEN, SERGEI KONYAGIN, AND TERENCE TAO

ABSTRACT. Let $G(X)$ denote the size of the largest gap between consecutive primes below X. Answering a question of Erdős, we show that

$$G(X) \geqslant f(X) \frac{\log X \log\log X \log\log\log\log X}{(\log\log\log X)^2},$$

where $f(X)$ is a function tending to infinity with X. Our proof combines existing arguments with a random construction covering a set of primes by arithmetic progressions. As such, we rely on recent work on the existence and distribution of long arithmetic progressions consisting entirely of primes.

CONTENTS

1. INTRODUCTION

Write $G(X)$ for the maximum gap between consecutive primes less than X. It is clear from the prime number theorem that

$$G(X) \geqslant (1 + o(1)) \log X,$$

as the *average* gap between the prime numbers which are $\leqslant X$ is $\sim \log X$. In 1931, Westzynthius [33] proved that infinitely often, the gap between consecutive prime numbers can be an arbitrarily large multiple of the average gap, that is, $G(X)/\log X \to \infty$ as $X \to \infty$. Moreover, he proved the qualitative bound[1]

$$G(X) \gg \frac{\log X \log_3 X}{\log_4 X}.$$

[1] As usual in the subject, $\log_2 x = \log\log x$, $\log_3 x = \log\log\log x$, and so on. The conventions for asymptotic notation such as $\ll$ and $o()$ will be defined in Section 1.2.

In 1935 Erdős [9] improved this to

$$G(X) \gg \frac{\log X \log_2 X}{(\log_3 X)^2}$$

and in 1938 Rankin [28] made a subsequent improvement

$$G(X) \geqslant (c+o(1)) \frac{\log X \log_2 X \log_4 X}{(\log_3 X)^2}$$

with $c = \frac{1}{3}$. The constant c was subsequently improved several times: to $\frac{1}{2}e^\gamma$ by Schönhage [30], then to $c = e^\gamma$ by Rankin [29], $c = 1.31256e^\gamma$ by Maier and Pomerance [24] and, most recently, $c = 2e^\gamma$ by Pintz [27].

Our aim in this paper is to show that c can be taken arbitrarily large.

Theorem 1. *Let $R > 0$. Then for any sufficiently large X, there are at least*

$$R \frac{\log X \log_2 X \log_4 X}{(\log_3 X)^2}$$

consecutive composite natural numbers not exceeding X.

In other words, we have

$$G(X) \geqslant f(X) \frac{\log X \log_2 X \log_4 X}{(\log_3 X)^2}$$

for some function $f(X)$ that goes to infinity as $X \to \infty$. Theorem 1 settles in the affirmative a long-standing conjecture of Erdős [10].

Theorem 1 has been simultaneously and independently established by Maynard [26] by a different method (relying on the sieve-theoretic techniques related to those used recently in [25] to obtain bounded gaps between primes, rather than results on linear equations between primes). As it turns out, the techniques of this paper and of that in [26] may be combined to establish further results on large prime gaps; see the followup paper [11] to this work and to [26].

Based on a probabilistic model of primes, Cramér [6] conjectured that[2]

$$\limsup_{X \to \infty} \frac{G(X)}{\log^2 X} = 1,$$

and Granville [14], using a refinement of Cramér's model, has conjectured that the $\limsup$ above is in fact at least $2e^{-\gamma} = 1.1229\ldots$. These conjectures are well beyond the reach of our methods. Cramér's model also predicts that the normalized prime gaps $\frac{p_{n+1}-p_n}{\log p_n}$ should have exponential distribution, that is, $p_{n+1} - p_n \geqslant C \log p_n$ for about $e^{-C}\pi(X)$ primes $\leqslant X$. Numerical evidence from prime calculations up to $4 \cdot 10^{18}$ [31] matches this prediction quite closely, with the exception of values of C close to $\log X$, in which there is very little data available. In fact, $\max_{X \leqslant 4 \cdot 10^{18}} G(X)/\log^2 X \approx 0.9206$, slightly below the predictions of Cramér and Granville.

Unconditional upper bounds for $G(X)$ are far from the conjectured truth, the best being $G(X) \ll X^{0.525}$ and due to Baker, Harman and Pintz [2]. Even the Riemann Hypothesis only[3] furnishes the bound $G(X) \ll X^{1/2} \log X$ [5].

[2] Cramér is not entirely explicit with this conjecture. In [6], he shows that his random analogues P_n of primes satisfy $\limsup (P_{n+1} - P_n)(\log P_n)^{-2} = 1$ and writes "Obviously we may take this as a suggestion that, for the particular sequence of ordinary prime numbers p_n, some similar relation may hold".

[3] Some slight improvements are available if one also assumes some form of the pair correlation conjecture; see [21].

All works on lower bounds for $G(X)$ have followed a similar overall plan of attack: show that there are at least $G(X)$ consecutive integers in $(X/2, X]$, each of which has a "very small" prime factor. To describe the results, we make the following definition.

Definition 1. *Let x be a positive integer. Define $Y(x)$ to be the largest integer y for which one may select residue classes $a_p \pmod p$, one for each prime $p \leqslant x$, which together "sieve out" (cover) the whole interval $[y] = \{1, \dots, y\}$.*

The relation between this function Y and gaps between primes is encoded in the following simple lemma.

Lemma 1.1. *Write $P(x)$ for the product of the primes less than or equal to x. Then we have $G(P(x) + Y(x) + x) \geqslant Y(x)$ for all x.*

Proof. Set $y = Y(x)$, and select residue classes $a_p \pmod p$, one for each prime $p \leqslant x$, which cover $[y]$. By the Chinese remainder theorem there is some m, $x < m \leqslant x + P(x)$, with $m \equiv -a_p \pmod p$ for all primes $p \leqslant x$. We claim that all of the numbers $m + 1, \dots, m + y$ are composite, which means that there is a gap of length y amongst the primes less than $m + y$, thereby concluding the proof of the lemma. To prove the claim, suppose that $1 \leqslant t \leqslant y$. Then there is some p such that $t \equiv a_p \pmod p$, and hence $m + t \equiv -a_p + a_p \equiv 0 \pmod p$, and thus p divides $m + t$. Since $m + t > m > x \geqslant p$, $m + t$ is indeed composite. □

By the prime number theorem we have $P(x) = e^{(1+o(1))x}$. It turns out (see below) that $Y(x)$ has size $x^{O(1)}$. Thus the bound of Lemma 1.1 implies that

$$G(X) \geqslant Y\big((1 + o(1)) \log X\big)$$

as $X \to \infty$. Theorem 1 follows from this and the following bound for Y, the proof of which is the main business of the paper.

Theorem 2. *For any $R > 0$ and for sufficiently large x we have*

$$Y(x) \geqslant R \frac{x \log x \log_3 x}{(\log_2 x)^2}. \tag{1.1}$$

The function Y is intimately related to *Jacobsthal's function* j. If n is a positive integer then $j(n)$ is defined to be the maximal gap between integers coprime to n. In particular $j(P(x))$ is the maximal gap between numbers free of prime factors $\leqslant x$, or equivalently 1 plus the longest string of consecutive integers, each divisible by some prime $p \leqslant x$. The construction given in the proof of Lemma 1.1 in fact proves that

$$j(P(x)) \geqslant Y\big((1 + o(1)) \log P(x)\big) = Y\big((1 + o(1))x\big).$$

This observation, together with results in the literature, gives upper bounds for Y. The best upper bound known is $Y(x) \ll x^2$, which comes from Iwaniec's work [23] on Jacobsthal's function. It is conjectured by Maier and Pomerance that in fact $Y(x) \ll x(\log x)^{2+o(1)}$. This places a serious (albeit conjectural) upper bound on how large gaps between primes we can hope to find via lower bounds for $Y(x)$: a bound in the region of $G(X) \gtrapprox \log X(\log \log X)^{2+o(1)}$, far from Cramér's conjecture, appears to be the absolute limit of such an approach.

We turn now to a discussion of the proof of Theorem 2. Recall that our task is to find y, as large as possible, so that the whole interval $[y]$ may be sieved using congruences $a_p \pmod p$, one for each prime $p \leqslant x$. Prior authors divided the sieving into different steps, a key to all of them being to take a common value of a_p for "large" p, say $a_p = 0$ for $z < p < \delta x$, where $\delta > 0$ is a small constant and $z = x^{c \log_3 x / \log_2 x}$

for some constant $c > 0$. The numbers in $[y]$ surviving this first sieving either have all of their prime factors $\leqslant z$ (i.e., they are "z-smooth") or are of the form pm with p prime and $m \leqslant y/\delta x$. One then appeals to bounds for smooth numbers, e.g. [3], to see that there are very few numbers of the first kind, say $O(x/\log^2 x)$. By the prime number theorem there are $\sim y \log_2 x/\log x$ unsieved numbers of the second kind. By contrast, if one were to take a random choice for a_p for $z < p < \delta x$, then with high probability, the number of unsifted integers in $[y]$ would be considerably larger, about $y \log z/\log x$.

One then performs a second sieving, choosing a_p for "small" $p \leqslant z$. Using a greedy algorithm, for instance, one can easily sieve out all but

$$\frac{y \log_2 x}{\log x} \prod_{p \leqslant z} \left(1 - \frac{1}{p}\right) \sim e^{-\gamma} \frac{y \log_2 x}{\log x \log z}$$

of the remaining numbers. There are alternative approaches using explicit choices for a_p; we will choose our a_p at random. (The set V of numbers surviving this second sieving has about the same size in each case.)

If $|V| \leqslant \pi(x) - \pi(\delta x)$, the number of "very large" primes, then we perform a (rather trivial) third sieving as follows: each $v \in V$ can be matched with one of these primes p, and one may simply take $a_p = v$. This is the route followed by all authors up to and including Rankin [29]; improvements to $G(x)$ up to this point depended on improved bounds for counts of smooth numbers. The new idea introduced by Maier and Pomerance [24] was to make the third sieving more efficient (and less trivial!) by using many $p \in (\delta x, x]$ to sift not one but *two* elements of V. To do this they established a kind of "twin primes on average" result implying that for most $p \in (\delta x, x]$, there are many pairs of elements of V that are congruent modulo p. Then the authors proved a crucial combinatorial result that *disjoint* sets V_p exist, each of two elements congruent modulo p, for a large proportion of these primes p; that is, for a large proportion of p, $a_p \mod p$ will sift out two elements of V, and the sifted elements are disjoint. Pintz [27] proved a "best possible" version of the combinatorial result, that in fact one can achieve a "nearly perfect matching", that is, disjoint sets V_p for almost all primes $p \in (\delta x, x]$, and this led to the heretofore best lower bound for $G(X)$.

Heuristically, much more along these lines should be possible. With y comparable to the right-hand side of (1.1), the set V turns out have expected cardinality comparable to a large multiple of $x/\log x$. Assuming that V is a "random" subset of $[y]$, for every prime $p \in (\delta x, x]$ there should in fact be a residue class $a \pmod p$ containing $\gg \log x/(\log_2 x)^{O(1)}$ elements of V. (Roughly, the heuristic predicts that the sizes of the sets $V \cap (a \pmod p)$ are Poisson distributed with parameter $\approx |V|/p$.) Whilst we cannot establish anything close to this, we are able to use almost all primes $p \in (x/2, x]$ to sieve r elements of V, for any fixed r. Where Maier and Pomerance appealed to (in fact proved) a result about pairs of primes on average, we use results about arithmetic progressions of primes of length r, established in work of the second and fourth authors [17], [16] and of these authors and Ziegler [19]. Specifically, we need results about progressions $q, q+r!p, q+2r!p, \ldots, q+(r-1)r!p$; if one ignores the technical factor $r!$, these are "progressions of primes with prime common difference". By taking $a_p = q$, the congruence $a_p \pmod p$ allows us to sift out all r elements of such a progression, and it is here that we proceed more efficiently than prior authors. Ensuring that many of these r-element sifted sets are disjoint (or at least have small intersections) is a rather difficult problem, however. Rather than dealing with these intersections directly, we utilize the random choice of a_p in the second step to prove that with high probability, V has a certain regularity with respect to intersections with progressions of the form $q, q + r!p, q + 2r!p, \ldots, q + (r-1)r!p$. We then prove that most elements of V survive the third sieving with uniformly small probability.

1.1. **Acknowledgments.** The research of SK was partially performed while he was visiting KF at the University of Illinois at Urbana–Champaign. Research of KF and SK was also carried out in part at the University of Chicago. KF and SK are thankful to Prof. Wilhelm Schlag for hosting these visits. KF also thanks the hospitality of the Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences.

Also, the research of the SK and TT was partially performed while SK was visiting the Institute for Pure and Applied Mathematics (IPAM) at UCLA, which is supported by the National Science Foundation.

KF was supported by NSF grant DMS-1201442. BG was supported by ERC Starting Grant 279438, *Approximate algebraic structure*. TT was supported by a Simons Investigator grant, the James and Carol Collins Chair, the Mathematical Analysis & Application Research Fund Endowment, and by NSF grant DMS-1266164.

Finally, all four of us wish to thank Andrew Granville and James Maynard for helpful discussions.

1.2. **Notational conventions.** We use $f = O(g)$ and $f \ll g$ to denote the claim that there is a constant $C > 0$ such that $|f(\cdot)| \leqslant Cg(\cdot)$ for all $\cdot$ in the domain of f. We adopt the convention that C is independent of any parameter unless such dependence is indicated by subscript such as $\ll_u$, except that C may depend on the parameter r (which we consider to be fixed) in Sections 2–4 and 6–7.

In Sections 2–4 and 6–7, the symbol $o(1)$ will stand for a function which tends to 0 as $x \to \infty$, uniform in all parameters except r unless otherwise indicated. The same convention applies to the asymptotic notation$f(x) \sim g(x)$, which means $f(x) = (1+o(1))g(x)$. In Sections 5 and the Appendix, $o(g(N))$ refers to some function $h(N)$ satisfying $\lim_{N\to\infty} h(N)/g(N) = 0$.

The symbols p, q and s will always denote prime numbers, except that in the the Appendix, s is a positive integer which measures the complexity of a system of linear forms.

Finally, we will be using the probabilistic method and will thus be working with finite probability spaces. Generically we write $\mathbb{P}$ for probability, and $\mathbb{E}$ for expectation. If a finite set A is equipped with the uniform probability measure, we write $\mathbb{P}_{a\in A}$ and $\mathbb{E}_{a\in A}$ for the associated probability and expectation. Variables in boldface will denote random real-valued scalars, while arrowed boldface symbols denote random vectors, e.g. $\vec{\mathbf{a}}$.

We also use $\#A$ to denote the cardinality of A, and for any positive real z, we let $[z] := \{n \in \mathbf{N} : 1 \leqslant n \leqslant z\}$ denote the set of natural numbers up to z.

2. On arithmetic progressions consisting of primes

A key tool in the proof of Theorem 2 is an asymptotic formula for counts of arithmetic progressions of primes. In fact, we shall be interested in progressions of primes of length r whose common difference is $r!$ times a prime[4], for positive integer values of r. The key technical result we shall need is Lemma 2.4 below. This is a relatively straightforward consequence of Lemma 2.1 below, which relies on the work on linear equations in primes of the second and fourth authors and Ziegler.

We turn to the details. Let y be a sufficiently large quantity (which goes to infinity for the purposes of asymptotic notation), and let x be a quantity that goes to infinity at a slightly slower rate than y; for sake of concreteness we will impose the hypotheses

$$x\sqrt{\log x} \leqslant y \leqslant x \log x. \tag{2.1}$$

[4] One could replace $r!$ here if desired by the slightly smaller *primorial* $P(r)$; as observed long ago by Lagrange and Waring [8], this primorial must divide the spacing of any sufficiently large arithmetic progression of primes of length r. However, replacing $r!$ by $P(r)$ would lead to only a negligible savings in the estimates here.

In fact the analysis in this section would apply under the slightly weaker hypotheses $y \log^{-O(1)} y \leqslant x \leqslant o(y)$, but we will stick with (2.1) for sake of concreteness since this condition will certainly be satisfied when applying the results of this section to prove Theorem 2. From (2.1) we see in particular that $\log y \sim \log x$, so we will use $\log x$ and $\log y$ more or less interchangeably in what follows. Let $\mathcal{P}$ denote the set of all primes in the interval $(x/2, x]$, and $\mathcal{Q}$ denote the set of all primes in the interval $(x/4, y]$; thus from the prime number theorem we have

$$\#\mathcal{P} \sim \frac{x}{2\log x}; \quad \#\mathcal{Q} \sim \frac{y}{\log x}. \tag{2.2}$$

In other words, $\mathcal{P}$ and $\mathcal{Q}$ both have density $\sim \frac{1}{\log x}$ inside $(x/2, x]$ and $(x/4, y]$ respectively.

Let $r \geqslant 1$ be a fixed natural number. We define a relation $\dashv$ between $\mathcal{P}$ and $\mathcal{Q}$ as follows: if $p \in \mathcal{P}$ and $q \in \mathcal{Q}$, we write $p \dashv q$ if the entire arithmetic progression $\{q, q + r!p, \dots, q + (r-1)r!p\}$ is contained inside $\mathcal{Q}$. One may think of the r relations $p \dashv q - ir!p$ for $i = 0, \dots, r-1$ as defining r different (but closely related) bipartite graphs between $\mathcal{P}$ and $\mathcal{Q}$. Note that if $p \dashv q$, then the residue class $q \pmod p$ is guaranteed to contain at least r primes from $\mathcal{Q}$, which is the main reason why we are interested in these relations (particularly for somewhat large values of r).

For our main argument, we will be interested in the typical degrees of the bipartite graphs associated to the relations $p \dashv q - ir!p$. Specifically, we are interested[5] in the following questions for a given $0 \leqslant i \leqslant r-1$:

(i) For a typical $p \in \mathcal{P}$, how many $q \in \mathcal{Q}$ are there such that $p \dashv q - ir!p$? (Note that the answer to this question does not depend on i.)
(ii) For a typical $q \in \mathcal{Q}$, how many $p \in \mathcal{P}$ are there such that $p \dashv q - ir!p$?

If $\mathcal{P}$ and $\mathcal{Q}$ were distributed randomly inside the intervals $(x/2, x]$ and $(x/4, y]$ respectively, with cardinalities given by (2.2), then standard probabilistic arguments (using for instance the Chernoff inequality) would suggest that the answer to question (i) is $\sim \frac{y}{\log^r x}$, while the answer to question (ii) is $\sim \frac{x}{2\log^r x}$. As it turns out, the local structure of the primes (for instance, the fact that all the elements of $\mathcal{P}$ and $\mathcal{Q}$ are coprime to $r!$) will bias the answers to each of these two questions; however (as one may expect from double counting considerations), they will be biased by exactly the same factor α_r (defined in (2.3) below), and the net effect of this bias will cancel itself out at the end of the proof of Theorem 2.

One can predict the answers to Questions (i) and (ii) using the Hardy-Littlewood prime tuples conjecture [22]. If we apply this conjecture (and ignore any issues as to how uniform the error term in that conjecture is with respect to various parameters), one soon arrives[6] at the prediction that the answer to Question (i) should be $\sim \alpha_r \frac{y}{\log^r x}$ for all $p \in \mathcal{P}$, and similarly the answer to Question (ii) should be $\sim \alpha_r \frac{x}{2\log^r x}$ for all $q \in \mathcal{Q}$, where for the rest of the paper α_r will denote the singular series

$$\alpha_r := \prod_{p \leqslant r} \left(\frac{p}{p-1}\right)^{r-1} \prod_{p > r} \frac{(p-r)p^{r-1}}{(p-1)^r}. \tag{2.3}$$

The exact form of α_r is not important for our argument, so long as it is finite, positive, and does not depend on x or y; but these claims are clear from (2.3) (note that the second factor $\frac{(p-r)p^{r-1}}{(p-1)^r}$ is non-zero and behaves asymptotically as $1 + O(1/p^2)$). As mentioned previously, this quantity will appear in two separate places in the proof of Theorem 2, but these two occurrences will eventually cancel each other out.

[5] Actually, for technical reasons we will eventually replace the relation $\dashv$ by slightly smaller relation $\dashv\!\!\|$, which will in turn be randomly refined to an even smaller relation $\overset{\vec{\mathbf{a}}}{\dashv\!\!\|}$; see below.

[6] See also Sections 6, 7 for some closely related computations.

The Hardy-Littlewood conjecture is still out of reach of current technology. Note that even the much weaker question as to whether the relation $p \dashv q$ is satisfied for at least *one* pair of p and q for any given r is at least as hard as establishing that the primes contain arbitrarily long arithmetic progressions, which was only established by the second and fourth authors in [15]. However, for the argument used to prove Theorem 2, it will suffice to be able to answer Question (i) for *almost all* $p \in \mathcal{P}$ rather than *all* $p \in \mathcal{P}$, and similarly for Question (ii). In other words, we only need (a special case of) the Hardy-Littlewood prime conjecture "on average". This is easier to establish; for instance, Balog [1] was able to use the circle method (or "linear Fourier analysis") to establish the prime tuples conjecture for "most" tuples in some sense. The results in [1] are not strong enough for our applications, because of our need to consider arbitrarily long arithmetic progressions (which are well-known to not be amenable to linear Fourier-analytic methods for $r \geqslant 4$, see [13]) rather than arbitrary prime tuples. Instead we will use (a modification of) the more recent work of the second and fourth authors [17]. More precisely, we claim the following bounds.

Lemma 2.1. *Let $x, y, r, \mathcal{P}, \mathcal{Q}$, and $\dashv$ be as above. Let $0 \leqslant i \leqslant r-1$.*

(i) *For all but $o(x/\log x)$ of the $p \in \mathcal{P}$, we have the estimate*

$$\#\{q \in \mathcal{Q} : p \dashv q - ir!p\} \sim \alpha_r \frac{y}{\log^r x}.$$

(ii) *For all but $o(y/\log x)$ of the $q \in \mathcal{Q}$, we have*

$$\#\{p \in \mathcal{P} : p \dashv q - ir!p\} \sim \alpha_r \frac{x}{2\log^r x}.$$

(iii) *For* all $p \in \mathcal{P}$*, we have the upper bounds*

$$\#\{q \in \mathcal{Q} : p \dashv q - ir!p\} \ll \frac{y}{\log^r x}.$$

(iv) *For* all $q \in \mathcal{Q}$*, we have the upper bounds*

$$\#\{p \in \mathcal{P} : p \dashv q - ir!p\} \ll \frac{x}{2\log^r x}.$$

Parts (iii) and (iv) follow from standard sieve-theoretic methods (e.g. the Selberg sieve); we omit the proof here, referring the reader instead[7] to [20] or [12]. The more interesting bounds are (i) and (ii). As stated above, these two claims are *almost* relatively straightforward consequences of the main result of the paper [17] of the second and fourth authors. However, some modifications of that work are required to deal with the fact that x and y are of somewhat different sizes. In Section 5 below we state the modified version of the main result of [17] that we need, Theorem 7. The deductions of parts (i) and (ii) of Lemmas 2.1 are rather similar to one another, and are given in Sections 6 and 7 respectively. Finally, a proof of Theorem 7 can be obtained by modifying the arguments of [17] in quite a straightforward manner, but in a large number of places. We record these modifications in Appendix A.

As presently defined, it is possible for the bipartite graphs given by the $p \dashv q - ir!p$ to overlap, thus it may happen that $p \dashv q - ir!p$ and $p \dashv q - jr!p$ for some $p \in \mathcal{P}$, $q \in \mathcal{Q}$, and $0 \leqslant i < j \leqslant r-1$. For instance, this situation will occur if $\mathcal{Q}$ has an arithmetic progression $q, q + r!p, \ldots, q + r \times r!p$ of length $r+1$ with $p \in \mathcal{P}$. For technical reasons, such overlaps are undesirable for our applications. However, these overlaps are rather rare and can be easily removed by the following simple device. We define the modified relation $\dashv\!\!|$ between $\mathcal{P}$ and $\mathcal{Q}$ by declaring $p \dashv\!\!| \, q$ if the progression $\{q, q + r!p, \ldots, q + (r-1)r!p\}$is contained inside $\mathcal{Q}$, but $q + r \times r!p$ does *not* lie in $\mathcal{Q}$. From construction we have the following basic fact:

[7] One could also deduce these bounds from Proposition 6.4' in Appendix A.

Lemma 2.2. *For any $p \in \mathcal{P}$ and $q \in \mathcal{Q}$ there is at most one $0 \leqslant i \leqslant r-1$ such that $p \dashv\!\mid q - ir!p$.*

We can then modify Lemma 2.1 slightly by replacing the relation $\dashv$ with its slightly perturbed version $\dashv\!\mid$:

Lemma 2.3. *Let $x, y, r, \mathcal{P}, \mathcal{Q}$, and $\dashv\!\mid$ be as above. Let $0 \leqslant i \leqslant r-1$.*

(i) *For all but $o(x/\log x)$ of the $p \in \mathcal{P}$, we have the estimate*

$$\#\{q \in \mathcal{Q} : p \dashv\!\mid q - ir!p\} \sim \alpha_r \frac{y}{\log^r x}.$$

(ii) *For all but $o(y/\log x)$ of the $q \in \mathcal{Q}$, we have*

$$\#\{p \in \mathcal{P} : p \dashv\!\mid q - ir!p\} \sim \alpha_r \frac{x}{2\log^r x}.$$

(iii) *For* all $p \in \mathcal{P}$, *we have the upper bounds*

$$\#\{q \in \mathcal{Q} : p \dashv\!\mid q - ir!p\} \ll \frac{y}{\log^r x}.$$

(iv) *For* all $q \in \mathcal{Q}$, *we have the upper bounds*

$$\#\{p \in \mathcal{P} : p \dashv\!\mid q - ir!p\} \ll \frac{x}{2\log^r x}.$$

Proof. Parts (iii) and (iv) are immediate from their counterparts in Lemma 2.1, since $\dashv\!\mid$ is a subrelation of $\dashv$. To prove (i), we simply observe from Lemma 2.1(iii) (with r replaced by $r+1$) that

$$\#\{q \in \mathcal{Q} : p \dashv q - ir!p \text{ but } p \not\dashv\!\mid q - ir!p\} \ll \frac{y}{\log^{r+1} x},$$

and the claim then follows from Lemma 2.1(i) and the triangle inequality. The claim (ii) is proven similarly. □

For technical reasons, it will be convenient to reformulate the main results of Lemma 2.3 as follows.

Lemma 2.4. *Let $x, y, r, \mathcal{P}, \mathcal{Q}$, and $\dashv\!\mid$ be as above. Then there exist subsets $\mathcal{P}_0$, $\mathcal{Q}_0$ of $\mathcal{P}$, $\mathcal{Q}$ respectively with*

$$\#\mathcal{P}_0 \sim \frac{x}{2\log x}; \quad \#\mathcal{Q}_0 \sim \frac{y}{\log x}, \tag{2.4}$$

such that

$$\#\{q \in \mathcal{Q} : p \dashv\!\mid q - ir!p\} \sim \alpha_r \frac{y}{\log^r x} \tag{2.5}$$

for all $p \in \mathcal{P}_0$ *and* $0 \leqslant i \leqslant r-1$, *and similarly that*

$$\#\{p \in \mathcal{P}_0 : p \dashv\!\mid q - ir!p\} \sim \alpha_r \frac{x}{2\log^r x} \tag{2.6}$$

for all $q \in \mathcal{Q}_0$ *and* $0 \leqslant i \leqslant r-1$.

Proof. From Lemma 2.3(i) we may already find a subset $\mathcal{P}_0$ of the desired cardinality obeying (2.5). If the $\mathcal{P}_0$ in (2.6) were replaced by $\mathcal{P}$, then a similar argument using Lemma 2.3(ii) (and taking the union bound for the exceptional sets for each $0 \leqslant i \leqslant r-1$) would give the remainder of the lemma. To deal with the presence of $\mathcal{P}_0$ in (2.6), it thus suffices to show that

$$\#\{p \in \mathcal{P} \backslash \mathcal{P}_0 : p \dashv\!\mid q - ir!p\} = o\left(\frac{x}{\log^r x}\right)$$

for all but $o(y/\log x)$ of the $q \in \mathcal{Q}$. By Markov's inequality, it suffices to show that

$$\#\{(p,q) \in (\mathcal{P}\backslash\mathcal{P}_0) \times \mathcal{Q} : p \dashv\!| \, q - ir!p\} = o\left(\frac{x}{\log^r x} \times \frac{y}{\log x}\right).$$

But this follows by summing Lemma 2.3(iii) for all $p \in \mathcal{P}\backslash\mathcal{P}_0$, since the set $\mathcal{P}\backslash\mathcal{P}_0$ has cardinality $o(x/\log x)$. □

3. Main construction

We now begin the proof of Theorem 2. It suffices to establish the following claim:

Theorem 3 (First reduction). *Let $r \geqslant 13$ be an integer. Take x to be sufficiently large depending on r (and going to infinity for the purposes of asymptotic notation), and then define y by the formula*

$$y := \frac{r}{6\log r}\frac{x\log x\log_3 x}{(\log_2 x)^2}. \tag{3.1}$$

Then there exists a residue class $a_s \pmod s$ for each prime $s \leqslant x$, such that the union of these classes contains every positive integer less than or equal to y.

The numerical values of 13 and 6 in the above theorem are only of minor significance, and can be ignored for a first reading.

Observe that x, y obey the condition (2.1) from the previous section. If Theorem 3 holds, then in terms of the quantity $Y(x)$ defined in the introduction, we have

$$Y(x) \geqslant y$$

which by (3.1) will imply Theorem 2 by taking r sufficiently large depending on R.

It remains to prove Theorem 3. Set

$$z := x^{\log_3 x/(3\log_2 x)}, \tag{3.2}$$

and partition the primes less than or equal to x into the four disjoint classes

$$\begin{aligned}
\mathcal{S}_1 &:= \{s \text{ prime} : s \leqslant \log x \text{ or } z < s \leqslant x/4\}\\
\mathcal{S}_2 &:= \{s \text{ prime} : \log x < s \leqslant z\}\\
\mathcal{S}_3 &:= \mathcal{P} = \{s \text{ prime} : x/2 < s \leqslant x\}\\
\mathcal{S}_4 &:= \{s \text{ prime} : x/4 < s \leqslant x/2\}.
\end{aligned}$$

We are going to sieve $[y]$ in four stages by removing at most one congruence class $a_s \pmod s$ for each prime $s \in S_i$, $i = 1, 2, 3, 4$. If we can do this in such a way that nothing is left at the end, we shall have achieved our goal.

We first dispose of the final sieving process (involving $\mathcal{S}_4$), as it is rather trivial. Namely, we reduce Theorem 3 to

Theorem 4 (Second reduction). *Let r, x, y be as in Theorem 3, and let $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$ be as above. Then there exists a residue class $a_s \pmod s$ for each $s \in \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$, such that the union of these classes contains all but at most $(\frac{1}{5} + o(1))\frac{x}{\log x}$ of the positive integers less than or equal to y.*

Indeed, if the $a_s \pmod{s}$ for $s \in \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$ are as in Theorem 4, then from the prime number theorem, the number of integers less than y that have not already been covered by a residue class is smaller than the number of primes in $\mathcal{S}_4$. Thus, we may eliminate each of these surviving integers using a residue class $a_s \pmod{s}$ from a different element s from $\mathcal{S}_4$ (and selecting residue classes arbitrarily for any $s \in \mathcal{S}_4$ that are left over), and Theorem 3 follows.

It remains to prove Theorem 4. For this, we perform the first sieving process (using up the primes from $\mathcal{S}_1$) and reduce to

Theorem 5 (Third reduction)**.** *Let $r, x, y, \mathcal{S}_2, \mathcal{S}_3$ be as in Theorem 4, and (as in the previous section) let $\mathcal{Q}$ denote the primes in the range $(x/4, y]$. Then there exists a residue class $a_s \pmod{s}$ for each $s \in \mathcal{S}_2 \cup \mathcal{S}_3$, such that the union of these classes contains all but at most $(\frac{1}{5} + o(1))\frac{x}{\log x}$ of the elements of $\mathcal{Q}$.*

Proof of Theorem 4 assuming Theorem 5. We take $a_s := 0$ for all $s \in \mathcal{S}_1$. Write $\mathcal{R} \subset [y]$ for the residual set of elements which survive this first sieving, that is to say $\mathcal{R}$ consists of all numbers in $[y]$ that are not divisible by any prime s in $\mathcal{S}_1$. Taking into account that $(x/4)\log x > y$ from (3.1), we conclude that

$$\mathcal{R} = \mathcal{Q} \cup \mathcal{R}^{\mathrm{err}},$$

where $\mathcal{R}^{\mathrm{err}}$ contains only z-smooth numbers, that is to say numbers in $[y]$ all of whose prime factors are at most z.

Let u denote the quantity

$$u := \frac{\log y}{\log z},$$

so from (3.2) one has $u \sim 3\frac{\log_2 x}{\log_3 x}$. By standard counts for smooth numbers (e.g. de Bruijn's theorem [3]),

$$\begin{aligned} \#\mathcal{R}^{\mathrm{err}} &\ll ye^{-u\log u + O(u \log\log(u+2))} \\ &= \frac{y}{\log^{3+o(1)} x} \\ &= \frac{x}{\log^{2+o(1)} x} \\ &= o(x/\log x). \end{aligned}$$

Thus the contribution of $\mathcal{R}^{\mathrm{err}}$ may be absorbed into the exceptional set in Theorem 4, and this theorem is now immediate from Theorem 5. □

Remark 1. One can replace the appeal to de Bruijn's theorem here by the simpler bounds of Rankin [28, Lemma II], if one makes the very minor change of increasing the 3 in the denominator of (3.2) to 4, and to similarly increase the 6 in (3.1) to 8.

It remains to establish Theorem 5. Recall from (2.2) that $\mathcal{Q}$ has cardinality $\sim y/\log x$. This is significantly larger than the error term of $(\frac{1}{5} + o(1))\frac{x}{\log x}$ permitted in Theorem 5; our sieving process has to reduce the size of $\mathcal{Q}$ by a factor comparable to y/x. The purpose of the second sieving, by congruences $\mathbf{a}_s \pmod{s}$ with $s \in \mathcal{S}_2$, is to achieve almost all of this size reduction. Our choice of the $\mathbf{a}_s$ for $s \in \mathcal{S}_2$ will be completely random (which is why we are using the boldface font here): that is, for each prime $s \in \mathcal{S}_2$ we select $\mathbf{a}_s$ uniformly at random from $\{0, 1, \ldots, s-1\}$, and these choices are independent for different values of s. Write $\vec{\mathbf{a}}$ for the random vector $(\mathbf{a}_s)_{s \in \mathcal{S}_2}$.

Observe that if n is any integer (not depending on $\vec{\mathbf{a}}$), then the probability that n lies outside of all of the $\mathbf{a}_s \pmod s$ is exactly equal to

$$\gamma := \prod_{s \in S_2} \left(1 - \frac{1}{s}\right).$$

This quantity will be an important normalizing factor in the arguments that follow. From Mertens' theorem and (3.1), (3.2) we see that

$$\gamma \sim \frac{\log_2 x}{\log z} \sim \frac{3(\log_2 x)^2}{\log x \log_3 x} \sim \frac{r}{2\log r}\frac{x}{y}. \tag{3.3}$$

Write $\mathcal{Q}(\vec{\mathbf{a}})$ for the (random) residual set of primes q in $\mathcal{Q}$ that do not lie in any of the congruence classes $\mathbf{a}_s \pmod s$ for $s \in \mathcal{S}_2$. We will in fact focus primarily on the slightly smaller set

$$\mathcal{Q}_0(\vec{\mathbf{a}}) := \mathcal{Q}(\vec{\mathbf{a}}) \cap \mathcal{Q}_0$$

where $\mathcal{Q}_0$ is the subset of $\mathcal{Q}$ constructed in Lemma 2.4. From linearity of expectation we see that

$$\mathbb{E}\#\mathcal{Q}(\vec{\mathbf{a}}) = \gamma\#\mathcal{Q} \tag{3.4}$$

and thus from (3.3), (2.2)

$$\mathbb{E}\#\mathcal{Q}(\vec{\mathbf{a}}) \sim \frac{r}{2\log r}\frac{x}{\log x}. \tag{3.5}$$

Similarly, from Lemma 2.4 we have

$$\#(\mathcal{Q}\backslash\mathcal{Q}_0) = o\left(\frac{y}{\log x}\right)$$

and thus from linearity of expectation and (3.3) we have

$$\mathbb{E}\#(\mathcal{Q}(\vec{\mathbf{a}})\backslash\mathcal{Q}_0(\vec{\mathbf{a}})) = o\left(\gamma\frac{y}{\log x}\right) = o\left(\frac{x}{\log x}\right).$$

In particular, from Markov's inequality we have

$$\#(\mathcal{Q}(\vec{\mathbf{a}})\backslash\mathcal{Q}_0(\vec{\mathbf{a}})) = o\left(\gamma\frac{y}{\log x}\right) = o\left(\frac{x}{\log x}\right) \tag{3.6}$$

with probability $1 - o(1)$.

We have an analogous concentration bound for $\#\mathcal{Q}(\vec{\mathbf{a}})$:

Lemma 3.1. *With probability* $1 - o(1)$*, we have*

$$\#\mathcal{Q}(\vec{\mathbf{a}}) \sim \frac{r}{2\log r}\frac{x}{\log x} \sim \gamma\frac{y}{\log x}.$$

In particular, from (3.6) *we also have*

$$\#\mathcal{Q}_0(\vec{\mathbf{a}}) \sim \frac{r}{2\log r}\frac{x}{\log x} \sim \gamma\frac{y}{\log x}$$

with probability $1 - o(1)$.

This lemma is proven by a routine application of the second moment method; we defer that proof to Section 4. It will now suffice to show

Set	Description	Expected cardinality
$\mathcal{P}$	Primes in $(x/2, x]$	$\sim \frac{x}{2\log x}$
$\mathcal{P}_0$	Primes in $\mathcal{P}$ connected to the expected # of primes in $\mathcal{Q}$	$\sim \frac{x}{2\log x}$
$\mathcal{P}_1(\vec{\mathbf{a}})$	Primes in $\mathcal{P}_0$ connected to the expected # of primes in $\mathcal{Q}(\vec{\mathbf{a}})$	$\sim \frac{x}{2\log x}$
$\mathcal{P}_1(\vec{\mathbf{a}}, q; i)$	Primes in $\mathcal{P}_1(\vec{\mathbf{a}})$ i-connected to a given prime $q \in \mathcal{Q}_1(\vec{\mathbf{a}})$	$\sim \gamma^{r-1}\alpha_r \frac{x}{2\log^r x}$
$\mathcal{Q}$	Primes in $(x/4, y]$	$\sim \frac{y}{\log x}$
$\mathcal{Q}_0$	Primes in $\mathcal{Q}$ connected to the expected # of primes in $\mathcal{P}_0$	$\sim \frac{y}{\log x}$
$\mathcal{Q}(\vec{\mathbf{a}})$	Randomly refined subset of $\mathcal{Q}$	$\sim \frac{r}{2\log r}\frac{x}{\log x}$
$\mathcal{Q}(\vec{\mathbf{a}}, p)$	Primes in $\mathcal{Q}(\vec{\mathbf{a}})$ connected to a given prime $p \in \mathcal{P}_1(\vec{\mathbf{a}})$	$\sim \gamma^r \alpha_r \frac{y}{\log^r x}$
$\mathcal{Q}_0(\vec{\mathbf{a}})$	Intersection of $\mathcal{Q}(\vec{\mathbf{a}})$ with $\mathcal{Q}_0$	$\sim \frac{r}{2\log r}\frac{x}{\log x}$
$\mathcal{Q}_1(\vec{\mathbf{a}})$	Primes in $\mathcal{Q}_0(\vec{\mathbf{a}})$ connected to the expected # of primes in $\mathcal{P}_1(\vec{\mathbf{a}})$	$\sim \frac{r}{2\log r}\frac{x}{\log x}$
$\mathcal{Q}_1(\vec{\mathbf{a}}, \vec{\mathbf{q}})$	Randomly refined subset of $\mathcal{Q}_1(\vec{\mathbf{a}})$	$\sim \frac{1}{2\log r}\frac{x}{\log x}$

TABLE 1. A brief description of the various $\mathcal{P}$ and $\mathcal{Q}$-type sets used in the construction, and their expected size. Roughly speaking, the congruence classes from $\mathcal{S}_1$ are used to cut down $[y]$ to approximately $\mathcal{Q}$, the congruence classes from $\mathcal{S}_2$ are used to cut $\mathcal{Q}$ down to approximately $\mathcal{Q}_0(\vec{\mathbf{a}})$, the congruence classes from $\mathcal{S}_3 = \mathcal{P}$ are used to cut $\mathcal{Q}_0(\vec{\mathbf{a}})$ down to approximately $\mathcal{Q}_1(\vec{\mathbf{a}}, \vec{\mathbf{q}})$, and the congruence classes in $\mathcal{S}_4$ are used to cover all surviving elements from previous sieving.

Theorem 6 (Fourth reduction)**.** *Let $x, y, r, \vec{\mathbf{a}}, \mathcal{P}_0, \mathcal{Q}_0(\vec{\mathbf{a}})$ be as above, and let $\varepsilon > 0$ be a quantity going to zero arbitrarily slowly as $x \to \infty$, thus $\varepsilon = o(1)$. Then with probability at least ε in the random choice of $\vec{\mathbf{a}}$, we may find a length r arithmetic progression $\{q_p + ir!p : 0 \leqslant i \leqslant r-1\}$ for each $p \in \mathcal{P}_0$, such that the union of these progressions contains all but at most $(\frac{1}{5} + o(1))\frac{x}{\log x}$ of the elements of $\mathcal{Q}_0(\vec{\mathbf{a}})$. (The $o(1)$ decay in the conclusion may depend on ε.)*

Indeed, from this theorem (and taking ε going to zero sufficiently slowly) we may find $\vec{\mathbf{a}}$ such that the conclusions of this theorem hold simultaneously with (3.6), and by combining the residue classes from $\vec{\mathbf{a}}$ with the residue classes $q_p \pmod p$ for $p \in \mathcal{P}_0$ from Theorem 6 (and selecting residue classes arbitrarily for $p \in \mathcal{P}\backslash\mathcal{P}_0$), we obtain Theorem 5.

It remains to establish Theorem 6. Note now (from Lemma 3.1) that we only need to reduce the size of the surviving set $\mathcal{Q}_0(\vec{\mathbf{a}})$ through sieving by a constant factor (comparable to $\frac{r}{\log r}$), rather than by a factor like y/x that goes to infinity as $x \to \infty$.

Recall from the previous section that we had the relation $\dashv\|$ between $\mathcal{P}$ and $\mathcal{Q}$. We now refine this relation to a (random) relation between $\mathcal{P}_0$ and $\mathcal{Q}(\vec{\mathbf{a}})$ as follows. If $p \in \mathcal{P}_0$ and $q \in \mathcal{Q}(\vec{\mathbf{a}})$, we write $p \overset{\vec{\mathbf{a}}}{\dashv\|} q$ if $p \dashv\| q$ and if the arithmetic progression $\{q, q + r!p, \ldots, q + (r-1)r!p\}$ is contained in $\mathcal{Q}(\vec{\mathbf{a}})$ (i.e. the entire progression survives the second sieving process).

Intuitively, if $p \in \mathcal{P}_0$ and $q \in \mathcal{Q}$ are such that $p \dashv\| q$, we expect $p \overset{\vec{\mathbf{a}}}{\dashv\|} q$ to occur with probability close to γ^r. The following lemma makes this intuition precise (compare with Lemma 2.4):

Lemma 3.2. *Let $\varepsilon > 0$ be a quantity going to zero arbitrarily slowly as $x \to \infty$. Then with probability at least ε, we can find (random) subsets $\mathcal{P}_1(\vec{\mathbf{a}})$ of $\mathcal{P}_0$ and $\mathcal{Q}_1(\vec{\mathbf{a}})$ of $\mathcal{Q}_0(\vec{\mathbf{a}})$ obeying the cardinality bounds*

$$\#\mathcal{P}_1(\vec{\mathbf{a}}) \sim \frac{x}{2\log x}; \quad \#\mathcal{Q}_1(\vec{\mathbf{a}}) \sim \#\mathcal{Q}_0(\vec{\mathbf{a}}) \sim \frac{r}{2\log r}\frac{x}{\log x}, \tag{3.7}$$

such that

$$\#\{q \in \mathcal{Q}(\vec{\mathbf{a}}) : p \overset{\vec{\mathbf{a}}}{\dashv\|} q - ir!p\} \sim \gamma^r \alpha_r \frac{y}{\log^r x}$$

for all $p \in \mathcal{P}_1(\vec{\mathbf{a}})$ *and* $0 \leqslant i \leqslant r-1$, *and such that*

$$\#\{p \in \mathcal{P}_1(\vec{\mathbf{a}}) : p \overset{\vec{\mathbf{a}}}{\dashv\|} q - ir!p\} \sim \gamma^{r-1} \alpha_r \frac{x}{2\log^r x}$$

for all $q \in \mathcal{Q}_1(\vec{\mathbf{a}})$ *and* $0 \leqslant i \leqslant r-1$. *(The implied* $o(1)$ *errors in the* $\sim$ *notation may depend on* ε*.)*

This lemma is also proven by an application of the second moment method; we defer this proof also to Section 4.

We are now ready to perform the third sieving process. Let us fix any $\vec{\mathbf{a}}$ obeying the properties in Lemma 3.2, and let $\mathcal{P}_1(\vec{\mathbf{a}})$ and $\mathcal{Q}_1(\vec{\mathbf{a}})$ be as in that lemma. Since $\vec{\mathbf{a}}$ has the desired properties with probability at least ε, in order to establish Theorem 6 (and thus Theorem 2 and Theorem 1), it suffices to show that for every such $\vec{\mathbf{a}}$, there is a choice of residue classes q_p for $p \in \mathcal{P}_0$ satisfying the required union property for Theorem 6.

For each $p \in \mathcal{P}_1(\vec{\mathbf{a}})$, we select $\mathbf{q}_p$ uniformly at random from the set

$$\mathcal{Q}(\vec{\mathbf{a}}, p) := \{q \in \mathcal{Q}(\vec{\mathbf{a}}) : p \overset{\vec{\mathbf{a}}}{\dashv\|} q\}, \tag{3.8}$$

with the $\mathbf{q}_p$ for $p \in \mathcal{P}_1(\vec{\mathbf{a}})$ being chosen independently (after $\vec{\mathbf{a}}$ has been fixed); note from Lemma 3.2 that

$$\#\mathcal{Q}(\vec{\mathbf{a}}, p) \sim \gamma^r \alpha_r \frac{y}{\log^r x} \tag{3.9}$$

for all $p \in \mathcal{P}_1(\vec{\mathbf{a}})$. We write $\vec{\mathbf{q}}$ for the random tuple $(\mathbf{q}_p)_{p \in \mathcal{P}_1(\vec{\mathbf{a}})}$, and for brevity write $\mathbb{P}_{\vec{\mathbf{q}}}$ and $\mathbb{E}_{\vec{\mathbf{q}}}$ for the associated probability and expectation with respect to this random tuple (where $\vec{\mathbf{a}}$ is now fixed). Let $\mathcal{Q}_1(\vec{\mathbf{a}}, \vec{\mathbf{q}})$ denote the elements of $\mathcal{Q}_1(\vec{\mathbf{a}})$ that are not covered by any of the arithmetic progressions $\{\mathbf{q}_p + ir!p : 0 \leqslant i \leqslant r-1\}$ for each $p \in \mathcal{P}_1(\vec{\mathbf{a}})$. We claim that

$$\mathbb{E}_{\vec{\mathbf{q}}} \#\mathcal{Q}_1(\vec{\mathbf{a}}, \vec{\mathbf{q}}) \leqslant \frac{x}{5\log x}. \tag{3.10}$$

This implies (for each fixed choice of $\vec{\mathbf{a}}$) the existence of a vector $\vec{q}$ with

$$\#\mathcal{Q}_1(\vec{\mathbf{a}}, \vec{q}) \leqslant \frac{x}{5\log x};$$

since $\#(\mathcal{Q}_0(\vec{\mathbf{a}}) \backslash \mathcal{Q}_1(\vec{\mathbf{a}})) = o(x/\log x)$ from (3.7), Theorem 6 follows (upon choosing q_p as the p component of $\vec{q}$ for $p \in \mathcal{P}_1(\vec{\mathbf{a}})$ and q_p arbitrarily for $p \in \mathcal{P}_0 \backslash \mathcal{P}_1(\vec{\mathbf{a}})$).

It remains to prove (3.10). We will shortly show that

$$\mathbb{P}_{\vec{\mathbf{q}}}(q \in \mathcal{Q}_1(\vec{\mathbf{a}}, \vec{\mathbf{q}})) \leqslant \frac{1+o(1)}{r} \tag{3.11}$$

for each $q \in \mathcal{Q}_1(\vec{\mathbf{a}})$. Assuming this bound, then from (3.7) and linearity of expectation we have

$$\mathbb{E}_{\vec{\mathbf{q}}} \#\mathcal{Q}_1(\vec{\mathbf{a}}, \vec{\mathbf{q}}) \leqslant \frac{1+o(1)}{r} \frac{r}{2\log r} \frac{x}{\log x}$$

which gives (3.10) as desired for $r \geqslant 13$.

It remains to prove (3.11). Fix $q \in \mathcal{Q}_1(\vec{\mathbf{a}})$, and consider the sets

$$\mathcal{P}_1(\vec{\mathbf{a}}, q; i) := \{p \in \mathcal{P}_1(\vec{\mathbf{a}}) : p \overset{\vec{\mathbf{a}}}{\dashv\|} q - ir!p\}$$

for $i = 0, \dots, r-1$. From Lemma 2.2, these sets are disjoint; from Lemma 3.2, these sets each have cardinality $(1+o(1))\gamma^{r-1}\alpha_r \frac{x}{2\log^r x}$.

Suppose that $0 \leqslant i \leqslant r-1$ and $p \in \mathcal{P}_1(\vec{\mathbf{a}}, q; i)$. Then $q - ir!p \in \mathcal{Q}(\vec{\mathbf{a}}, p)$ by (3.8), and the probability that $\mathbf{q}_p = q - ir!p$ is equal to

$$\frac{1}{\#\mathcal{Q}(\vec{\mathbf{a}}, p)} = \frac{1+o(1)}{\gamma^r \alpha_r \frac{y}{\log^r x}}$$

thanks to (3.9). By independence, the probability that $\mathbf{q}_p \neq q - ir!p$ for all $0 \leqslant i \leqslant r-1$ and $p \in \mathcal{P}_1(\vec{\mathbf{a}}, q; i)$ (which is a necessary condition for q to end up in $\mathcal{Q}_1(\vec{\mathbf{a}}, \vec{\mathbf{q}})$) is thus

$$\begin{aligned}\prod_{i=0}^{r-1} \prod_{p \in \mathcal{P}_1(\vec{\mathbf{a}}, q; i)} \left(1 - \frac{1+o(1)}{\gamma^r \alpha_r \frac{y}{\log^r x}}\right) &= \exp\left(-\frac{1+o(1)}{\gamma^r \alpha_r \frac{y}{\log^r x}} r(1+o(1))\gamma^{r-1}\alpha_r \frac{x}{2\log^r x}\right) \\ &= \exp\left(-(1+o(1))\frac{rx}{2y\gamma}\right) \\ &= \exp(-(1+o(1))\log r)\end{aligned}$$

by (3.3). The claim (3.11) follows.

4. Probability estimates

In this section we establish the results left unproven in the last section, namely Lemmas 3.1 and 3.2. Our primary tool here will be the second moment method. Throughout, the probabilistic quantities we write are all with respect to the random choice of the vector $\vec{\mathbf{a}} = (\mathbf{a}_s)_{s \in S_2}$. In several of these proofs we will make use of the quantities γ_i defined by

$$\gamma_i := \prod_{s \in \mathcal{S}_2} \left(1 - \frac{i}{s}\right) \tag{4.1}$$

for $i = 1, \dots, 2r$. Note that $\gamma_1 = \gamma$ in the notation of the previous section.

Lemma 4.1. *We have $\gamma_i \sim \gamma^i$, uniformly for all $1 \leqslant i \leqslant 2r$.*

Proof. We have, uniformly for $1 \leqslant i \leqslant 2r$,

$$\gamma_i = \gamma^i \prod_{s \in \mathcal{S}_2} \left(1 - \frac{i}{s}\right)\left(1 - \frac{1}{s}\right)^{-i} = \gamma^i \prod_{s \in \mathcal{S}_2} \left(1 + O(s^{-2})\right) = \gamma^i(1 + O(1/\log x)),$$

using the fact that all primes $s \in \mathcal{S}_2$ are $> \log x$. □

4.1. **Proof of Lemma 3.1.** To prove Lemma 3.1 we use the second moment method. Indeed, from Chebyshev's inequality it will suffice to prove the asymptotics

$$\mathbb{E}\#\mathcal{Q}(\vec{\mathbf{a}}) \sim \gamma \frac{y}{\log x} \tag{4.2}$$

and

$$\mathbb{E}(\#\mathcal{Q}(\vec{\mathbf{a}}))^2 \sim \left(\gamma \frac{y}{\log x}\right)^2. \tag{4.3}$$

The claim (4.2) is just (3.5), so we turn to (4.3). The left-hand side of (4.3) may be written as

$$\sum_{q_1,q_2\in\mathcal{Q}} \mathbb{P}(q_1,q_2 \in \mathcal{Q}(\vec{\mathbf{a}})).$$

The diagonal contribution $q_1 = q_2$ is clearly negligible (it is crudely bounded by $\#\mathcal{Q}$, which is much smaller than $\left(\gamma\frac{y}{\log x}\right)^2$), so by (2.2) it suffices to show that

$$\mathbb{P}(q_1,q_2 \in \mathcal{Q}(\vec{\mathbf{a}})) \sim \gamma^2$$

for any *distinct* $q_1, q_2 \in \mathcal{Q}$.

Fix any such q_1, q_2. Observe that for each $s \in \mathcal{S}_2$, the probability that q_1 and q_2 simultaneously avoid $\mathbf{a}_s \pmod s$ is equal to $1-\frac{2}{s}$ if s does not divide $q_2 - q_1$, and $1-\frac{1}{s}$ otherwise. In the latter case, we crudely write $1-\frac{1}{s}$ as $(1+O(\frac{1}{\log x}))(1-\frac{2}{s})$. Since $q_2 - q_1 = O(y)$ and all the primes in $\mathcal{S}_2$ are at least $\log x$, we see that there are at most $O(\frac{\log y}{\log\log x}) = o(\log x)$ primes s that divide $q_2 - q_1$. We conclude that

$$\mathbb{P}(q_1,q_2 \in \mathcal{Q}(\vec{\mathbf{a}})) = \left(1+O\left(\frac{1}{\log x}\right)\right)^{o(\log x)} \prod_{s\in\mathcal{S}_2}\left(1-\frac{2}{s}\right) \sim \gamma_2,$$

and the claim now follows from Lemma 4.1.

4.2. **A preliminary lemma.** In order to establish Lemma 3.2, we will first need the following preliminary result in this direction.

Lemma 4.2. *The following two claims hold with probability* $1-o(1)$ *(in the random choice of* $\vec{\mathbf{a}}$*), and for any* $0 \leqslant i \leqslant r-1$.

(i) *One has*

$$\#\{q \in \mathcal{Q}(\vec{\mathbf{a}}) : p \overset{\vec{\mathbf{a}}}{\dashv\|} q - ir!p\} \sim \gamma^r \alpha_r \frac{y}{\log^r x} \sim \gamma^r \#\{q \in \mathcal{Q} : p \dashv\| q - ir!p\} \tag{4.4}$$

for all but $o(x/\log x)$ *values of* $p \in \mathcal{P}_0$.

(ii) *One has*

$$\#\{p \in \mathcal{P}_0 : p \overset{\vec{\mathbf{a}}}{\dashv\|} q - ir!p\} \sim \gamma^{r-1} \alpha_r \frac{x}{2\log^r x} \sim \gamma^{r-1} \#\{p \in \mathcal{P}_0 : p \dashv\| q - ir!p\} \tag{4.5}$$

for all but $o(x/\log x)$ *values of* $q \in \mathcal{Q}_0(\vec{\mathbf{a}})$.

We begin with the proof of Lemma 4.2(i), which goes along very similar lines to that of the previous lemma. As the quantities here do not depend on i, we may take $i = 0$. The second part of (4.4) follows from (2.5), so it suffices to show that with probability $1-o(1)$, we have

$$\#\{q \in \mathcal{Q}(\vec{\mathbf{a}}) : p \overset{\vec{\mathbf{a}}}{\dashv\|} q\} \sim \gamma^r \alpha_r \frac{y}{\log^r x} \tag{4.6}$$

for all but $o(x/\log x)$ values of $p \in \mathcal{P}_0$. By Markov's inequality and (2.4), it suffices to show that for each $p \in \mathcal{P}_0$, we have the event (4.6) with probability $1-o(1)$.

Fix $p \in \mathcal{P}_0$. By Chebyshev's inequality, it suffices to show that

$$\mathbb{E}\#\{q \in \mathcal{Q}(\vec{\mathbf{a}}) : p \overset{\vec{\mathbf{a}}}{\dashv\|} q\} \sim \gamma^r \alpha_r \frac{y}{\log^r x}$$

and

$$\mathbb{E}\big(\#\{q \in \mathcal{Q}(\vec{\mathbf{a}}) : p \overset{\vec{\mathbf{a}}}{\dashv\|} q\}\big)^2 \sim \left(\gamma^r \alpha_r \frac{y}{\log^r x}\right)^2.$$

By (2.5), Lemma 4.1, and linearity of expectation, it thus suffices to show that

$$\mathbb{P}(q, q + r!p, \dots, q + (r-1)r!p \in \mathcal{Q}(\vec{\mathbf{a}})) \sim \gamma_r \tag{4.7}$$

for all $q \in \mathcal{Q}$ with $p \dashv\| q$, and similarly that

$$\mathbb{P}(q_1, q_1 + r!p, \dots, q_1 + (r-1)r!p, q_2, q_2 + r!p, \dots, q_2 + (r-1)r!p \in \mathcal{Q}(\vec{\mathbf{a}})) \sim \gamma_{2r} \tag{4.8}$$

for any distinct $q_1, q_2 \in \mathcal{Q}$ with $p \dashv\| q_1, p \dashv\| q_2$.

We begin with (4.7). For any $s \in \mathcal{S}_2$, the probability that $q, q + r!p, \dots, q + (r-1)r!p$ simultaneously avoid $\mathbf{a}_s \pmod{s}$ is equal to $1 - \frac{r}{s}$ (note that s is coprime to $r!p$). So (4.7) then follows (with exact equality) from (4.1) and independence.

Now we turn to (4.8). For any $s \in \mathcal{S}_2$, the probability that $q_1, q_1 + r!p, \dots, q_1 + (r-1)r!p, q_2, q_2 + r!p, \dots, q_2 + (r-1)r!p$ simultaneously avoid $\mathbf{a}_s \pmod{s}$ is usually equal to $1 - \frac{2r}{s}$; the exceptions arise when s divides $q_2 - q_1 + ir!p$ for some $-r \leqslant i \leqslant r$, in which case the probability is instead $(1 + O(\frac{1}{\log x}))(1 - \frac{2r}{s})$. But by arguing as in the proof of Lemma 3.1, the number of exceptional s is $o(\log x)$. Multiplying all the independent probabilities together, we obtain the claim (4.8). This concludes the proof of Lemma 4.2(i).

Now we prove Lemma 4.2(ii). Again, the second part of (4.5) follows from (2.6). For the first part, it suffices (by Lemma 3.1 and (3.3)) to show that with probability $1 - o(1)$, one has

$$\sum_{q \in \mathcal{Q}_0(\vec{\mathbf{a}})} \left| \#\{p \in \mathcal{P}_0 : p \overset{\vec{\mathbf{a}}}{\dashv\|} q - ir!p\} - \gamma^{r-1}\alpha_r \frac{x}{2\log^r x} \right|^2 = o\left(\gamma \frac{y}{\log x}\left(\gamma^{r-1}\frac{x}{\log^r x}\right)^2\right).$$

By Markov's inequality, it suffices to show that

$$\mathbb{E} \sum_{q \in \mathcal{Q}_0(\vec{\mathbf{a}})} \left| \#\{p \in \mathcal{P}_0 : p \overset{\vec{\mathbf{a}}}{\dashv\|} q - ir!p\} - \gamma^{r-1}\alpha_r \frac{x}{2\log^r x} \right|^2 = o\left(\gamma \frac{y}{\log x}\left(\gamma^{r-1}\frac{x}{\log^r x}\right)^2\right).$$

Expanding out the square, it suffices to show the estimate

$$\mathbb{E} \sum_{q \in \mathcal{Q}_0(\vec{\mathbf{a}})} \left(\#\{p \in \mathcal{P}_0 : p \overset{\vec{\mathbf{a}}}{\dashv\|} q - ir!p\}\right)^b \sim \gamma \frac{y}{\log x}\left(\gamma^{r-1}\alpha_r \frac{x}{2\log^r x}\right)^b \tag{4.9}$$

for $b = 0, 1, 2$.

The $b = 0$ case of (4.9) follows from (2.2) and (3.4). For the $b = 1, 2$ cases, observe from Lemma 2.4 that

$$\sum_{q \in \mathcal{Q}_0} (\#\{p \in \mathcal{P}_0 : p \dashv\| q - ir!p\})^b \sim \left(\alpha_r \frac{x}{2\log^r x}\right)^b \frac{y}{\log x}.$$

By linearity of expectation, it thus suffices to show that

$$\mathbb{P}(q - ir!p, q + (1-i)r!p, \dots, q + (r-1-i)r!p \in \mathcal{Q}_0(\vec{\mathbf{a}})) \sim \gamma^r \tag{4.10}$$

whenever $p \in \mathcal{P}_0$, $q \in \mathcal{Q}_0$ with $p \dashv\| q - ir!p$, and

$$\mathbb{P}(q + jr!p_k \in \mathcal{Q}_0(\vec{\mathbf{a}}) \text{ for all } j = -i, 1-i, \dots, r-1-i \text{ and } k = 1, 2) \sim \gamma^{2r-1} \tag{4.11}$$

whenever $p_1, p_2 \in \mathcal{P}_0$, $q \in \mathcal{Q}_0$ with $p_1 \dashv\!\| \, q - ir!p_1$, $p_2 \dashv\!\| \, q - ir!p_2$, and $p_1 \neq p_2$ (the total contribution of the diagonal $p_1 = p_2$ is easily seen to be negligible).

We begin with the proof of (4.10). For any $s \in \mathcal{S}_2$, the probability that the progression $q - ir!p, q + (1 - i)r!p, \ldots, q + (r - 1 - i)r!p$ avoids $\mathbf{a}_s \pmod s$ is equal to $1 - \frac{r}{s}$ (since s is coprime to $r!p$), and so by (4.1) and independence the left-hand side of (4.10) is precisely γ_r. The claim now follows from Lemma 4.1.

Now we prove (4.11). For any $s \in \mathcal{S}_2$, the probability that the intersecting progressions $q - ir!p_1, q + (1 - i)r!p_1, \ldots, q + (r - 1 - i)r!p_1$ and $q - ir!p_2, q + (1 - i)r!p_2, \ldots, q + (r - 1 - i)r!p_2$ avoid s is usually $1 - \frac{2r-1}{s}$ (note that q is a common value of the two arithmetic progressions). The exceptions occur when s divides $jp_1 + kp_2$ for some $-r \leqslant j, k \leqslant r$ that are not both zero, but by arguing as before we see that the number of such exceptions is $o(\log x)$, and the probability in these cases is $(1 + O(\frac{1}{\log x}))(1 - \frac{2r-1}{s})$. Thus by independence, the left-hand of (4.11) is $\sim \gamma_{2r-1}$, and the claim follows from Lemma 4.1. The proof of Lemma 4.2 is now complete.

4.3. **Proof of Lemma 3.2.** Suppose that $\varepsilon > 0$ goes to zero as $x \to \infty$ sufficiently slowly.

Let $\mathcal{P}_1(\vec{\mathbf{a}})$ be the set of $p \in \mathcal{P}_0$ obeying (4.4) for all $0 \leqslant i \leqslant r - 1$ (actually the choice of i is irrelevant here), then from Lemma 4.2(i) and (2.4) we have that with probability at least $1 - \varepsilon$ we have

$$\#\mathcal{P}_1(\vec{\mathbf{a}}) \sim \frac{x}{2\log x} \tag{4.12}$$

as required. From Lemma 3.1 we also have $\#\mathcal{Q}_0(\vec{\mathbf{a}}) \sim \frac{r}{2\log r}\frac{x}{\log x}$ with probability at least $1 - \varepsilon$ as required. To finish the proof of the lemma, it suffices in view of Lemma 4.2(ii) to show that with probability at least 3ε, one has

$$\#\{p \in \mathcal{P}_0 \backslash \mathcal{P}_1(\vec{\mathbf{a}}) : p \overset{\vec{\mathbf{a}}}{\dashv\!\|} \, q - ir!p\} = o(\gamma^{r-1} x / \log^r x)$$

for all but $o(x/\log x)$ values of $q \in \mathcal{Q}_0(\vec{\mathbf{a}})$, and any $0 \leqslant i \leqslant r - 1$.

We use a double counting argument. It clearly suffices to show with probability at least 3ε that

$$\#\{(p, q) \in (\mathcal{P}_0 \backslash \mathcal{P}_1(\vec{\mathbf{a}})) \times \mathcal{Q}_0(\vec{\mathbf{a}}) : p \overset{\vec{\mathbf{a}}}{\dashv\!\|} \, q - ir!p\} = o\left(\gamma^{r-1}\frac{x}{\log^r x} \times \frac{x}{\log x}\right)$$

for all $0 \leqslant i \leqslant r - 1$. Actually, the left-hand side does not depend on i (as can be seen by shifting q by $ir!p$), so it suffices to show that the above holds with $i = 0$. By (3.3), we may rewrite this requirement as

$$\#\{(p, q) \in (\mathcal{P}_0 \backslash \mathcal{P}_1(\vec{\mathbf{a}})) \times \mathcal{Q}_0(\vec{\mathbf{a}}) : p \overset{\vec{\mathbf{a}}}{\dashv\!\|} \, q\} = o\left(\gamma^r \alpha_r \frac{y}{\log^r x} \times \frac{x}{\log x}\right).$$

Now from (4.4) and (4.12) we have

$$\#\{(p, q) \in \mathcal{P}_1(\vec{\mathbf{a}}) \times \mathcal{Q}_0(\vec{\mathbf{a}}) : p \overset{\vec{\mathbf{a}}}{\dashv\!\|} \, q\} \sim \gamma^r \alpha_r \frac{y}{\log^r x} \times \frac{x}{2\log x}$$

with probability at least $1 - \varepsilon$, so it suffices to show that

$$\#\{(p, q) \in \mathcal{P}_0 \times \mathcal{Q}_0(\vec{\mathbf{a}}) : p \overset{\vec{\mathbf{a}}}{\dashv\!\|} \, q\} \leqslant \frac{1 + o(1)}{1 - 4\varepsilon}\gamma^r \alpha_r \frac{y}{\log^r x} \times \frac{x}{2\log x}$$

with probability at least 4ε (recall that $\varepsilon = o(1)$). By Markov's inequality, it thus suffices to show that

$$\mathbb{E}\#\{(p, q) \in \mathcal{P}_0 \times \mathcal{Q}_0(\vec{\mathbf{a}}) : p \overset{\vec{\mathbf{a}}}{\dashv\!\|} \, q\} \leqslant (1 + o(1))\gamma^r \alpha_r \frac{y}{\log^r x} \times \frac{x}{2\log x}.$$

But this follows from the $b = 1$ case of (4.9). The proof of Lemma 3.2 is now complete.

5. LINEAR EQUATIONS IN PRIMES WITH LARGE SHIFTS

The paper [17] of the second and fourth author is concerned with counting the number of prime points parameterized by a system of affine-linear forms in a convex body, when the constant terms in the affine-linear forms are comparable to the size of the body. To establish Lemma 2.1 we will require a strengthening of the main result in [17], in which the constant terms in the affine-linear forms are permitted to be larger than the size of the body by a logarithmic factor. The aim of this section is to state this strengthening. The proof involves a number of minor modifications to the arguments of [17]: these are indicated in Appendix A.

To state the results, we need to recall some notation from [17]. If $d, t \geqslant 1$ be integers, then an *affine-linear form* on $\mathbb{Z}^d$ is a function $\psi : \mathbb{Z}^d \to \mathbb{Z}$ which is the sum $\psi = \dot{\psi} + \psi(0)$ of a homogeneous linear form $\dot{\psi} : \mathbb{Z}^d \to \mathbb{Z}$ and a constant $\psi(0) \in \mathbb{Z}$. A *system of affine-linear forms* on $\mathbb{Z}^d$ is a collection $\Psi = (\psi_1, \dots, \psi_t)$ of affine-linear forms on $\mathbb{Z}^d$. A system Ψ is said to have finite complexity if and only if no form $\dot{\psi}_i$ is a multiple of any other form $\dot{\psi}_j$.

We recall that the *von Mangoldt function* $\Lambda(n)$ is defined to equal $\log p$ when n is a prime p or a power of that prime, and zero otherwise.

Here is the main result of [17].

Theorem A [17, Main Theorem]. *Let N, d, t, L be positive integers, and let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of affine-linear forms of finite complexity with*

$$\|\Psi\|_N \leqslant L. \tag{5.1}$$

Let $K \subset [-N, N]^d$ be a convex body. Then we have

$$\sum_{\vec{n} \in K \cap \mathbb{Z}^d} \prod_{i=1}^{t} \Lambda(\psi_i(\vec{n})) = \beta_\infty \prod_p \beta_p + o_{t,d,L}(N^d) \tag{5.2}$$

where

$$\beta_\infty := \mathrm{vol}_d \left(K \cap \Psi^{-1}((\mathbb{R}^+)^t) \right)$$

and

$$\beta_p := \mathbb{E}_{\vec{n} \in (\mathbb{Z}/p\mathbb{Z})^d} \prod_{i=1}^{t} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(\vec{n})).$$

Here $\|\Psi\|_N$ is defined by

$$\|\Psi\|_N := \sum_{i=1}^{t} \sum_{j=1}^{d} |\dot{\psi}_i(e_j)| + \sum_{i=1}^{t} \left| \frac{\psi_i(0)}{N} \right|.$$

The function $\Lambda_{\mathbb{Z}/p\mathbb{Z}} : \mathbb{Z} \to \mathbb{R}^+$ is the local von Mangoldt function, *that is the p-periodic function defined by setting $\Lambda_{\mathbb{Z}/p\mathbb{Z}}(b) := \frac{p}{p-1}$ when b is coprime to p and $\Lambda_{\mathbb{Z}/p\mathbb{Z}}(b) = 0$ otherwise. Also, $\{e_1, \dots, e_d\}$ is the standard basis for $\mathbb{R}^d$.*

Strictly speaking, the results in [17] were conditional on two (at the time unproven) conjectures, namely the Möbius-Nilsequences conjecture and the inverse conjecture for the Gowers uniformity norms. However, these conjectures have since been proven in [16] and [19] respectively, and so the above theorem is now unconditional.

The variant of this result that we shall need is that in which the condition (5.1) is replaced by the weaker condition

$$\|\Psi\|_{N,B} \leqslant L, \tag{5.3}$$

where $B > 0$ is some constant (in fact any $B > 1$ will suffice for us). Here we have defined

$$\|\Psi\|_{N,B} := \sum_{i=1}^{t} \sum_{j=1}^{d} |\dot{\psi}_i(e_j)| + \sum_{i=1}^{t} \left| \frac{\psi_i(0)}{N \log^B N} \right|.$$

Note that $\|\Psi\|_{N,0} = \|\Psi\|_N$.

The conclusion is the same, except that the error term in (5.2) must also depend on B.

Theorem 7. *Let $B > 0$ be a positive quantity. Let everything be as in Theorem A, except assume that instead of condition* (5.1) *we have only the weaker condition* (5.3)*. Then we have*

$$\sum_{\vec{n} \in K \cap \mathbb{Z}^d} \prod_{i=1}^{t} \Lambda(\psi_i(\vec{n})) = \beta_\infty \prod_p \beta_p + o_{t,d,L,B}(N^d),$$

where β_∞ and the β_p are given by the same formulae as before.

This extension in effect allows us to consider affine linear forms in which the constant terms $\psi_i(0)$ can have size up to $\asymp N \log^B N$, whereas in Theorem A, they are restricted to have size $O(N)$. As mentioned above, the proof of Theorem 7 is deferred to Appendix A.

6. Proof of Lemma 2.1(i)

In this section we deduce Lemma 2.1(i) from Theorem 7. Throughout this section, x and y obey (2.1), all $o(1)$ terms may depend on r, and α_r is defined in (2.3).

It suffices to prove the lemma when x is an integer, which we henceforth assume. We first partition the range $(x/4, y]$ of q into blocks of size about x, so that p and q range over intervals of roughly the same size. Namely, for a non-negative integer m and $u \in \mathbb{R}$ we write

$$I(m, u) := \mathbb{Z} \cap [mx, (m+1)x) \cap (x/4, \infty) \cap [0, y - r!(r-1)u].$$

Observe that

$$\sum_{0 \leqslant m \leqslant y/x} \#I(m, n_1) \sim y \tag{6.1}$$

uniformly for $x/2 < n_1 \leqslant x$ and that

$$\#(m, n_1) = x \quad \text{for all } x/2 < n_1 \leqslant x \tag{6.2}$$

for all except $o(y/x)$ values of m, $0 \leqslant m \leqslant y/x$. We call these exceptional values of m *bad* and the remaining $0 \leqslant m \leqslant y/x$ obeying (6.2) *good*. Trivially $|I(m, n_1)| \leqslant x$ for all m, n_1.

We claim the following estimate:

Proposition 1. *We have*

$$\sum_{\substack{0 \leqslant m \leqslant y/x \\ x/2 < n_1 \leqslant x}} |F(m, n_1)|^2 \Lambda(n_1) = o(yx^2) \tag{6.3}$$

where

$$F(m,n_1) := \sum_{n_2 \in I(m,n_1)} \left(\prod_{j=0}^{r-1} \Lambda(n_2 + jr!n_1) - \alpha_r \right). \tag{6.4}$$

Let us assume this proposition for the moment and conclude the proof of Lemma 2.1(i). Let $\varepsilon = \varepsilon(x) > 0$ with ε decaying to zero sufficiently slowly. If n_1 is a prime in $(x/2, x]$, say that n_1 is *exceptional* and write $n_1 \in \mathscr{E}$ if the number of q for which $x/4 < q < y - (r-1)r!n_1$ and $q + jr!n_1$ is prime for $j = 0, \dots, r-1$ differs from $\alpha_r y/\log^r x$ by at least $\varepsilon y/\log^r x$. It follows straightforwardly that if $n_1 \in \mathscr{E}$ then

$$\left| \sum_{x/4 \leqslant n_2 < y-(r-1)r!n_1} \prod_{j=0}^{r-1} \Lambda(n_2 + jr!n_1) - \alpha_r y \right| \geqslant \frac{1}{2}\varepsilon y$$

if x is sufficiently large. (To see this, note that due to the restriction on the ranges of n_1, n_2, $\Lambda(n_2 + jr!n_1) = \log x + O(\log_2 x)$ whenever $n_2 + jr!n_1$ is prime. Λ is also supported on prime powers, but the contribution from these is negligible.) Recall the definition (6.4) of $F(m,n_1)$. Using the fact that $[x/4, y-(r-1)r!n_1] = \bigcup_m I(m,n_1)$ and (6.1), we conclude that

$$\left| \sum_{0 \leqslant m \leqslant y/x} F(m,n_1) \right| \geqslant \frac{1}{4}\varepsilon y$$

for sufficiently large x. By Cauchy's inequality, we thus have

$$\sum_{0 \leqslant m \leqslant y/x} |F(m,n_1)|^2 \geqslant \frac{\left(\frac{1}{4}\varepsilon y\right)^2}{\frac{y}{x} + 2} \geqslant \frac{1}{32}\varepsilon^2 xy \qquad (n_1 \in \mathscr{E}).$$

Since $\Lambda(n_1) = \log n_1 \geqslant \log(x/2)$ for every prime n_1, we therefore see that the left-hand side of (6.3) is at least

$$\frac{1}{32}\varepsilon^2 xy \log(x/2) \# \mathscr{E}.$$

Applying (6.3), we conclude that $\#\mathscr{E} = o(x/\log x)$ if ε goes to zero slowly enough, and Lemma 2.1(i) follows.

We now prove the proposition. After a change of variables, the left-hand side of (6.3) may be written as

$$\sum_{\substack{0 \leqslant m \leqslant y/x \\ x/2 < n_1 \leqslant x}} \left| \sum_{n_2 \in I(m,n_1) - mx} \left(\prod_{j=0}^{r-1} \Lambda(n_2 + jr!n_1 + mx) - \alpha_r \right) \right|^2 \Lambda(n_1).$$

Expanding out the square, we can write this expression as

$$\sum_{0 \leqslant m \leqslant y/x} \Sigma_2(m) - 2\alpha_r \sum_{0 \leqslant m \leqslant y/x} \Sigma_1(m) + \alpha_r^2 \sum_{0 \leqslant m \leqslant y/x} \Sigma_0(m)$$

where $\Sigma_2(m), \Sigma_1(m), \Sigma_0(m)$ are the quantities

$$\Sigma_2(m) := \sum_{\substack{x/2<n_1\leqslant x \\ n_2\in I(m,n_1)-mx \\ n_3\in I(m,n_1)-mx}} \Lambda(n_1) \prod_{\substack{0\leqslant j\leqslant r-1 \\ \ell=2,3}} \Lambda(n_\ell + jr!n_1 + mx)$$

$$\Sigma_1(m) := \sum_{\substack{x/2<n_1\leqslant x \\ n_2\in I(m,n_1)-mx}} (\#I(m,n_1))\Lambda(n_1) \prod_{j=0}^{r-1} \Lambda(n_2 + jr!n_1 + mx)$$

$$\Sigma_0(m) := \sum_{x/2<n_1\leqslant x} (\#I(m,n_1))^2\Lambda(n_1).$$

To prove (6.3), it will thus suffice to establish the estimates

$$\sum_{0\leqslant m\leqslant y/x} \Sigma_b(m) \sim \alpha_r^b \frac{yx^2}{2} \tag{6.5}$$

for $b = 0, 1, 2$.

We begin with the $b = 2$ case, which is the most difficult. We apply Theorem 7 with $d := 3$, $t := 2r + 1$, and the forms $\Psi = (\psi_1, \dots, \psi_{2r+1})$ given by

$$\Psi(n_1, n_2, n_3) := (n_1, (n_\ell + jr!n_1 + mx)_{0\leqslant j\leqslant r-1, \ell=2,3})$$

and convex polytope $K = K(m)$ given by

$$K(m) := \{(u_1, u_2, u_3) \in \mathbb{R}^3 : x/2 < u_1 \leqslant x, u_2, u_3 \in I(m, u_1) - mx\}.$$

Since $\Psi(K(m)) \subset (\mathbb{R}^+)^{2r+1}$, it follows from Theorem 7 that

$$\Sigma_2(m) = \mathrm{vol}(K(m)) \prod_p \beta_p + o(x^3), \tag{6.6}$$

where

$$\beta_p := \mathbb{E}_{\vec{n}\in(\mathbb{Z}/p\mathbb{Z})^3} \prod_{i=1}^{2r+1} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(\vec{n})).$$

Obviously the system Ψ has finite complexity.

We claim that

$$\beta_p = \begin{cases} (\frac{p}{p-1})^{2(r-1)} & p \leqslant r \\ \left(\frac{(p-r)p^{r-1}}{(p-1)^r}\right)^2 & p > r. \end{cases} \tag{6.7}$$

The proof of the claim is quite straightforward. Indeed if $p \leqslant r$ then, modulo p, $n_2 + jr!n_1 + mx \equiv n_2 + mx$ and $n_3 + jr!n_1 + mx \equiv n_3 + mx$, and so all the forms $\psi_i(\vec{n})$ are coprime to p if and only if none of $n_1, n_2 + mx$ or $n_3 + mx$ is zero mod p. Thus the number of $\vec{n} = (n_1, n_2, n_3)$ for which all of the forms $\psi_i(\vec{n})$ are nonzero mod p is precisely $(p - 1)^3$.

If, by contrast, $p > r$ then either $n_1 \equiv 0 \pmod p$ or else the values of $n_2 + jr!n_1 + mx$, $0 \leqslant j < r$ are all distinct mod p, and hence at most one of them can be zero. The same is true for the values of $n_3 + jr!n_1 + mx$. Thus if $n_1 \not\equiv 0 \pmod p$ then there are r values of n_2 for which one of the forms $\psi_i(\vec{n})$ vanishes, and also r values of n_3 for which one of these forms vanishes, and thus $2rp - r^2$ pairs (n_2, n_3) in total. Thus in this case the number of $\vec{n} = (n_1, n_2, n_3)$ for which all of the forms $\psi_i(\vec{n})$ are nonzero mod p is $p^3 - p^2 - (p-1)(2rp - r^2) = (p-1)(p-r)^2$, and this confirms the formula for β_p.

It follows from the claim (6.7) and the definition (2.3) of α_r that $\prod_p \beta_p = \alpha_r^2$ and hence, by (6.6), that

$$\Sigma_2(m) = \mathrm{vol}(K(m))\alpha_r^2 + o(x^3).$$

By (6.2) above we have $\mathrm{vol}(K(m)) = x^3/2$ for all good values of m, and $\mathrm{vol}(K(m)) \leqslant x^3/2$ for all m. It is thus straightforward to conclude the required asymptotic (6.5) for $b = 2$.

Next we turn to the $b = 1$ case of (6.5). Define

$$S_1(m) := \sum_{\substack{x/2 < n_1 \leqslant x \\ 0 \leqslant n_2 < x}} \Lambda(n_1) \prod_{j=0}^{r-1} \Lambda(n_2 + jr!n_1 + mx).$$

Then, by (6.2),

$$x \sum_{m \text{ good}} S_1(m) \leqslant \sum_m \Sigma_1(m) \leqslant x \sum_{0 \leqslant m \leqslant y/x} S_1(m). \tag{6.8}$$

To estimate $S_1(m)$, apply Theorem 7 with $d := 2$, $t := r+1$, forms $\Psi = (\psi_1, \dots, \psi_{r+1})$ given by

$$\Psi(n_1, n_2) := (n_1, (n_2 + jr!n_1 + mx)_{0 \leqslant j < r})$$

and convex polytope $K := (x/2, x] \times [0, x)$. The system Ψ also has finite complexity. Noting that $\Psi(K) \subset (\mathbb{R}^+)^{r+1}$, we obtain

$$S_1(m) = \frac{x^2}{2} \prod_p \beta_p + o(x^2) \tag{6.9}$$

uniformly in m where

$$\beta_p := \mathbb{E}_{\vec{n} \in (\mathbb{Z}/p\mathbb{Z})^2} \prod_{i=1}^{r+1} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(\vec{n})).$$

We claim that

$$\beta_p = \begin{cases} (\frac{p}{p-1})^{r-1} & p \leqslant r \\ \frac{(p-r)p^{r-1}}{(p-1)^r} & p > r. \end{cases} \tag{6.10}$$

The proof of the claim is similar to that of (6.7) but rather easier. Indeed if $p \leqslant r$ then, modulo p, $n_2 + jr!n_1 + mx \equiv n_2 + mx$, and so all the forms $\psi_i(\vec{n})$ are coprime to p if and only if neither n_1 nor $n_2 + mx$ is zero mod p, and so the number of $\vec{n} = (n_1, n_2)$ for which all of the forms $\psi_i(\vec{n})$ are nonzero mod p is precisely $(p-1)^2$.

If $p > r$ then either $n_1 \equiv 0 \pmod p$ or else the values of $n_2 + jr!n_1 + mx$, $0 \leqslant j < r$ are all distinct mod p, and hence at most one of them can be zero. Thus if $n_1 \not\equiv 0 \pmod p$ then there are r values of n_2 for which one of the forms $\psi_i(\vec{n})$ vanishes. Thus in this case the number of $\vec{n} = (n_1, n_2)$ for which all of the forms $\psi_i(\vec{n})$ are nonzero mod p is $p^2 - p - (p-1)r = (p-1)(p-r)$, and this confirms the formula for β_p.

From (6.10) and (2.3) we have $\prod_p \beta_p = \alpha_r$. It follows from (6.8), (6.9) and (6.10) that (6.5) holds for $b = 1$.

Finally we establish the $b = 0$ case of (6.5). By (6.2), for all except $o(y/x)$ bad values of m we have $\#I(m, n_1) = x$. If m is good then by the prime number theorem $\Sigma_0(m) \sim x^3/2$, and so the contribution to $\sum_m \Sigma_0(m)$ from the good m is $\sim yx^2/2$. The contribution from the bad m can be absorbed into the error term, and so (6.5) for $b = 0$ follows. The proof of Lemma 2.1(i) is now complete.

7. PROOF OF LEMMA 2.1(II)

In this section we deduce Lemma 2.1(ii) from Theorem 7. The argument is very similar to that in the last section. As before, x and y obey (2.1), all $o(1)$ terms may depend on r, and α_r is defined in (2.3).

We may again assume that x is an integer. The analogue of Proposition 1 is

Proposition 2. *We have*

$$\sum_{x/4<n_1\leqslant y} |F(n_1)|^2\Lambda(n_1) = o(yx^2) \tag{7.1}$$

where

$$F(n_1) := \sum_{x/2<n_2\leqslant x} \left(\Lambda(n_2) \prod_{\substack{-i\leqslant j<r-i \\ j\neq 0}} \Lambda(n_1 + jr!n_2) - \alpha_r \right). \tag{7.2}$$

Let us assume this proposition for the moment and conclude the proof of Lemma 2.1(ii). Let $\varepsilon = \varepsilon(x) > 0$ tend to 0 as $x \to \infty$ sufficiently slowly. If n_1 is a prime in $(x/4, y]$, we say that n_1 is *exceptional* and write $n_1 \in \mathscr{E}$ if the number of primes $p \in (x/2, x]$ for which $n_1 + jr!p$ is a prime in $(x/4, y]$ for $-i \leqslant j < r - i$, $j \neq 0$, differs from $\alpha_r(x/2)/\log^r x$ by at least $\varepsilon x/\log^r x$. Arguing as in the proof of Lemma 2.1(i), if $n_1 \in \mathscr{E}$ then for sufficently large x we have

$$\left| \sum_{\substack{x/2<n_2\leqslant x \\ n_1-ir!n_2>x/4 \\ n_1+(r-i-1)r!n_2\leqslant y}} \prod_{\substack{-i\leqslant j<r-i \\ j\neq 0}} \Lambda(n_1 + jr!n_2) - \frac{1}{2}\alpha_r x \right| \geqslant \frac{1}{2}\varepsilon x. \tag{7.3}$$

Note that the second and third conditions in the summation are precisely what constrain all the $n_1 + jr!n_2$, $-i \leqslant j < r - i$, to lie in $(x/4, y]$. If we assume that

$$(r + 1)!x < n_1 < y - (r + 1)!x$$

and recall from (7.2) above the definition of $F(n_1)$, we see that (7.3) is equivalent to

$$|F(n_1)| \geqslant \frac{1}{2}\varepsilon x.$$

Since $\Lambda(n_1) = \log n_1 \geqslant \log(x/4)$ for all prime n_1, we conclude from the prime number theorem that the left-hand side of (7.1) is at least

$$\left(\frac{1}{2}\varepsilon x\right)^2 \log(x/4)\big(\#\mathscr{E} - O(x/\log x)\big).$$

From this and (7.1) we conclude that $\#\mathscr{E} = o(y/\log x)$ provided ε tends to zero sufficiently slowly, and Lemma 2.1(ii) follows.

It remains to establish Proposition 2. For this, we break up the range of n_1 as in the proof of Lemma 2.1 (i). For a non-negative integer m define

$$I(m) := \mathbb{Z} \cap [mx, (m+1)x) \cap (x/4, y].$$

Then we may decompose the left-hand side of (7.1) as

$$\sum_{\substack{0 \leqslant m \leqslant y/x \\ n_1 \in I(m)}} |F(n_1)|^2 \Lambda(n_1),$$

With a simple change of variables we see that this quantity equals

$$\sum_{\substack{0 \leqslant m \leqslant y/x \\ n_1 \in I(m)-mx}} \left| \sum_{x/2 < n_2 \leqslant x} \Lambda(n_2) \prod_{\substack{-i \leqslant j < r-i \\ j \neq 0}} \Lambda(n_1 + jr! n_2 + mx) - \frac{1}{2}\alpha_r x \right|^2 \Lambda(n_1 + mx).$$

Expanding out the square and applying the prime number theorem, we may therefore express the above quantity as

$$\sum_{0 \leqslant m \leqslant y/x} \Sigma_2(m) - \alpha_r x \sum_{0 \leqslant m \leqslant y/x} \Sigma_1(m) + \frac{1}{4}\alpha_r^2 x^2 y + o(x^2 y),$$

where

$$\Sigma_2(m) := \sum_{\substack{n_1 \in I(m)-mx \\ x/2 < n_2 \leqslant x \\ x/2 < n_3 \leqslant x}} \Lambda(n_1 + mx)\Lambda(n_2)\Lambda(n_3) \prod_{\substack{-i \leqslant j < r-i \\ j \neq 0}} \prod_{\ell = 2,3} \Lambda(n_1 + jr! n_\ell + mx)$$

and

$$\Sigma_1(m) = \sum_{\substack{n_1 \in I(m)-mx \\ x/2 < n_2 \leqslant x}} \Lambda(n_2) \prod_{-i \leqslant j < r-i} \Lambda(n_1 + jr! n_2 + mx).$$

It will thus suffice to show that

$$\sum_{0 \leqslant m \leqslant y/x} \Sigma_b(m) \sim \left(\alpha_r \frac{x}{2}\right)^b y \tag{7.4}$$

for $b = 1, 2$.

We first handle the $b = 2$ case of (7.4). We can estimate $\Sigma_2(m)$ using Theorem 7 with $d := 3$, $t := 2r+1$, forms $\Psi = (\psi_1, \dots, \psi_{2r+1})$ given by

$$\Psi(n_1, n_2, n_3) := (n_1 + mx, n_2, n_3, (n_1 + jr! n_\ell + mx)_{-i \leqslant j < r-i, j \neq 0, \ell = 2,3})$$

and convex polytope $K(m) := (I(m) - mx) \times (x/2, x] \times (x/2, x]$. The theorem tells us that uniformly in m we have

$$\Sigma_2(m) = \mathrm{vol}(K(m)) \prod_p \beta_p + o(x^3), \tag{7.5}$$

where again

$$\beta_p := \mathbb{E}_{\vec{n} \in (\mathbb{Z}/p\mathbb{Z})^3} \prod_{i=1}^{2r+1} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(\vec{n})).$$

It is again clear that the system Ψ has finite complexity.

Now we claim that the β_p are given by the same formulae as in (6.7), that is to say

$$\beta_p = \begin{cases} (\frac{p}{p-1})^{2(r-1)} & p \leqslant r \\ \left(\frac{(p-r)p^{r-1}}{(p-1)^r}\right)^2 & p > r. \end{cases}$$

The proof of this is very similar to that of (6.7), but subtly different. If $p \leqslant r$ then the forms $\psi_i(\vec{n})$ are all equal to one of $n_1 + mx, n_2, n_3$ mod p, and so there are $(p-1)^3$ choices of $\vec{n} \in (\mathbb{Z}/p\mathbb{Z})^3$ for which all the forms are coprime to p. If $p > r$ then we must choose $n_1 \not\equiv -mx \pmod{p}$. For any such choice there are precisely r choices of n_2 for which one of $n_1 + jr!n_2 + mx$ $(-i \leqslant j < r-i, j \neq 0)$ and n_2 is $0 \pmod{p}$, namely $n_2 \equiv 0 \pmod{p}$ and $n_2 \equiv -(jr!)^{-1}(n_1+mx) \pmod{p}$ for $-i \leqslant j < r-i, j \neq 0$. Similarly there are precisely r choices for which one of $n_1 + jr!n_3 + mx$ $(-i \leqslant j < r-i, j \neq 0)$ and n_3 is $0 \pmod{p}$, and so we have $2rp - r^2$ bad choices of (n_2, n_3) for each $n_1 \not\equiv -mx \pmod{p}$. Therefore, as before, the number of choices of $\vec{n}$ for which at least one of the $\psi_i(\vec{n})$ vanishes mod p is $p^3 - p^2 - (2rp - p^2) = (p-1)(p-r)^2$.

Therefore $\prod_p \beta_p = \alpha_r^2$, and hence from (7.5) we have

$$\sum_{0 \leqslant m \leqslant y/x} \Sigma_2(m) = \alpha_r^2 \sum_{0 \leqslant m \leqslant y/x} \mathrm{vol}(K(m)) + o(yx^2).$$

We have $\#I(m) = x$ and hence $\mathrm{vol}(K(m)) = x^3/4$ for all except $o(y/x)$ values of m, and so the $b = 2$ case of (7.4) follows immediately.

Now we turn our attention to the $b = 1$ case of (7.4). Again we can estimate it using Theorem 7, now with $d := 2$, $t := r + 1$, forms $\Psi = (\psi_1, \dots, \psi_{r+1})$ given by

$$\Psi(n_1, n_2) := (n_2, (n_1 + jr!n_2 + mx)_{-i \leqslant j < r-i})$$

and convex polytope $K(m) := (I(m) - mx) \times (x/2, x]$.

The theorem tells us that uniformly in m we have

$$\Sigma_1(m) = \mathrm{vol}(K(m)) \prod_p \beta_p + o(x^2), \tag{7.6}$$

where

$$\beta_p := \mathbb{E}_{\vec{n} \in (\mathbb{Z}/p\mathbb{Z})^2} \prod_{i=1}^{r+1} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(\vec{n})).$$

Again the system Ψ has finite complexity.

We claim that the β_p are the same as in (6.10), that is to say

$$\beta_p = \begin{cases} (\frac{p}{p-1})^{r-1} & p \leqslant r \\ \frac{(p-r)p^{r-1}}{(p-1)^r} & p > r. \end{cases}$$

Indeed if $p \leqslant r$ then, mod p, all the forms in Ψ are either $n_1 + mx$ or n_2, so there are $(p-1)^2$ choices of $\vec{n}$ for which all of the $\psi_i(\vec{n})$ are coprime to p. If $p > r$ then we must take $n_1 \not\equiv -mx \pmod{p}$, and then for each such choice there are precisely r values of $n_2 \pmod{p}$ for which one of the $\psi_i(\vec{n})$ is $0 \pmod{p}$, namely $n_2 \equiv 0 \pmod{p}$ and $n_2 \equiv -(jr!)^{-1}(n_1 + mx) \pmod{p}$ for $-i \leqslant j < r - i$, $j \neq 0$. It follows that there are $p^2 - p - r(p-1) = (p-1)(p-r)$ choices of $\vec{n} \in (\mathbb{Z}/p\mathbb{Z})^2$ for which none of the $\psi_i(\vec{n})$ is 0 $\pmod{p}$.

Therefore $\prod_p \beta_p = \alpha_r$, and hence from (7.6) we have

$$\sum_{0\leqslant m\leqslant y/x} \Sigma_1(m) = \alpha_r \sum_{0\leqslant m\leqslant y/x} \mathrm{vol}(K(m)) + o(yx).$$

We have $\#I(m) = x$ and hence $\mathrm{vol}(K(m)) = x^2/2$ for all except $o(y/x)$ values of m, and so the $b = 1$ case of (7.4) follows immediately. The proof of Lemma 2.1(ii) is now complete.

8. Further comments and speculations

The reduction of Theorem 5 to Theorem 6 was somewhat wasteful, as one replaced the entire residue class $q_p \pmod p$ by a fairly short arithmetic progression $q_p, q_p + r!p, \dots, q_p + (r-1)r!p$ inside that residue class. One could attempt to strengthen the argument here by working with more general patterns such as $q_p, q_p + a_1 r!p, \dots, q_p + a_r r!p$ for some $0 < a_1 < \cdots < a_r = o(y/x)$, and possibly trying to exploit further averaging over the $a_1, \dots, a_r$. However, we were unable to take advantage of such ideas to make any noticeable improvements to the arguments or results.

The dependence of R on x in Theorem 1 is completely ineffective, for two different reasons. The sources of this ineffectivity are

- The use of Davenport's ineffective bound
$$\sup_\theta |\mathbb{E}_{n\in[N]}\mu(n)e(n\theta)| \ll_A \log^{-A} N$$
in [16], which is intimately related to the possibility of Siegel zeros; and
- the use of ultrafilter arguments in [19] (and in other work of the inverse conjectures for the Gowers norms, such as that of Szegedy [32]).

The first source of ineffectivity appears to be less serious than the second with our present state of knowledge. For example, if one is only interested in having the conclusion of Theorem 1 for an infinite sequence of x's (rather than all sufficiently large x) then by choosing x judiciously the influence of Siegel zeros can be avoided and one has an effective version of Davenport's bound. See [7] for some related discussion.

The second source of ineffectivity is problematic, since in taking R large we need inverse theorems for the Gowers $U^{s+1}[N]$-norm with $s = s(R)$ tending to infinity. Proofs not using ultrafilters are only known in the cases $s = 2, 3$ and 4, and the bounds in the inverse theorem [18] for the Gowers $U^4[N]$-norm (which were not worked out in that paper) are already incredibly bad, of "$\log_*$ type" or worse. In principle (but with great pain) the ultrafilters in [19] could be removed, but the bounds would be similarly bad. It seems that a genuinely new idea is needed to make these bounds, and thus the approach of the present paper, effective in any reasonable sense (for example R being bounded below by $\log_k x$ for some finite k).

Appendix A. Linear equations in primes

In this appendix all page numbers refer to the published version of the paper [17]*, with which we assume a certain familiarity.*

We turn now to the proof of Theorem 7, indicating the points at which we must be careful assuming only the bound $\|\Psi\|_{N,B} \leqslant L$ rather than the stronger bound $\|\Psi\|_N \leqslant L$ allowed in Theorem A, which is the main theorem of [17]. The key points are that (a) the sieve-theoretic portions of [17] are essentially unaffected by shifts, and (b) the Möbius-nilsequences conjecture used in [17] comes with a savings of $\log^{-A} N$ for arbitrary $A > 0$, which is enough to absorb the effect of shifting for that portion of the argument.

We require a precise measure of the *complexity* of the system Ψ (cf. [17, Definition 1.5]) which plays a crucial role in the arguments. If $1 \leqslant i \leqslant t$ and $s \geqslant 0$ then we say that Ψ has i-complexity at most s if one can cover the $t-1$ forms $\{\dot{\psi}_j : j \neq i\}$ by $s+1$ classes, such that $\dot{\psi}_i$ does not lie in the linear span of any of these classes. The *complexity* of the system of forms Ψ is defined to be the least s for which the system has i-complexity at most s for all $1 \leqslant i \leqslant t$, or ∞ if no such s exists. Note that a system Ψ has finite complexity if and only if no form $\dot{\psi}_i$ is a multiple of any other form $\dot{\psi}_j$.

Let us first of all note that [17] was written to be conditional upon two sets of conjectures, the Möbius and Nilsequences Conjecture MN(s) and the Inverse Conjectures for the Gowers norms GI(s) which were unproven at the time in the cases $s \geqslant 3$. These are now theorems, established in [16] and [19] respectively, and thus the results of [17] which we plan to modify in this section are unconditional. We have no need to change any aspect of the inner workings of either [16] or [19].

The argument in [17] proceeds via a series of reductions to other statements. First, in [17, Chapter 4], some straightforward linear algebra reductions are given. The first part of the chapter concerns [17, Theorem 1.8] and does not concern us here; our interest begins near the top of page 1771. The subsection "*Elimination of the archimedean factor*" makes no use of any bound on $\|\Psi\|_N$. This section allows us to assume henceforth that $\psi_1, \dots, \psi_t > N^{9/10}$ on K. The only change we need to make to the next subsection, "*Normal form reduction of the main theorem*" is to replace $\|\cdot\|_N$ in the statement of [17, Lemma 4.4] by $\|\cdot\|_{N,B}$. That such a variant is valid follows from the proof of [17, Lemma 4.4] and in particular the observation that $\tilde{\Psi}(0) = \Psi(0)$, where $\tilde{\Psi} : \mathbb{Z}^{d'} \to \mathbb{Z}^t$ is the system of forms constructed in that proof.

The rest of [17, Chapter 4] carries over unchanged. Thus (changing L to $\tilde{L} = O_{d,t,L}(1)$) we may assume henceforth that our system affine-linear forms ψ_i is in s-normal form and still satisfies $\|\Psi\|_{N,B} \leqslant L$.

The next step, undertaken in [17, Chapter 5] is to decompose the sum

$$\sum_{\vec{n} \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} \Lambda(\psi_i(\vec{n}))$$

along progressions with common difference $W = \prod_{p \leqslant w} p$, where $w = \log\log\log N$ (say). This is the "W-trick". The task of proving Theorem A is reduced to that of establishing the estimate ([17, Theorem 5.1])

$$\sum_{\vec{n} \in K \cap \mathbb{Z}^d} \left(\prod_{i \in [t]} \Lambda'_{b_i,W}(\psi_i(\vec{n})) - 1 \right) = o(N^d)$$

with $b_1, \dots, b_t \in [W]$ coprime to W, uniformly in the choice of b_i. Here

$$\Lambda'_{b,W}(n) := \frac{\phi(W)}{W} \Lambda'(Wn + b)$$

and Λ' denotes the restriction of Λ to the primes.

We claim that the proof of Theorem 7 may be similarly reduced to the task of establishing

$$\sum_{\vec{n} \in K \cap \mathbb{Z}^d} \left(\prod_{i \in [t]} \Lambda'_{b_i,W}(\psi_i(\vec{n})) - 1 \right) = o_B(N^d) \tag{A.1}$$

uniformly for $b_1, \dots, b_t \in [W]$, but now only assuming the weaker condition $\|\Psi\|_{N,B} \leqslant L$.

The reduction proceeds exactly as in [17, Chapter 5], except that at the bottom of page 1777 we must remark that the constant term $\tilde{\psi}_{i,a}(0)$ is now only bounded by $O_{L,d,t}(N\log^B N/W)$, and where on page 1778 we said that $\|\tilde{\Psi}\|_{\tilde{N}} = O(1)$, we must now say that $\|\tilde{\Psi}\|_{\tilde{N},B} = O(1)$.

The desired estimate (A.1) may be written in the equivalent form

$$\sum_{\vec{n}\in K\cap\mathbb{Z}^d}\left(\prod_{i\in[t]} T^{\psi_i(0)}\Lambda'_{b_i,W}(\dot{\psi}_i(\vec{n})) - 1\right) = o_B(N^d), \tag{A.2}$$

where $\dot{\psi}_i$ denotes the homogeneous (linear) part of the affine form ψ_i and T denotes the translation operator defined by $T^a f(x) := f(x+a)$. The homogeneous system $\dot{\Psi} = (\dot{\psi}_1,\dots,\dot{\psi}_t)$ satisfies the condition $\|\tilde{\Psi}\|_N \leqslant L$.

The first step in proving (A.2) is to prove a variant of [17, Proposition 6.4] for the shifted functions $T^{\psi_i(0)}\Lambda'_{b_i,W}$. We claim that in fact the following generalisation of that proposition holds (for notation and further discussion, see [17, Chapter 6]).

Proposition 6.4'. *Let $D > 1$ be arbitrary, and let $z_1,\dots,z_t \in \mathbb{Z}_{\geqslant 0}$, $z_i \leqslant N^{1.01}$, be arbitrary shifts. Then there is a constant $C_0 := C_0(D)$ such that the following is true. Let $C \geqslant C_0$, and suppose that $N' \in [CN, 2CN]$. Let $b_1,\dots,b_t \in \{0,1,\dots,W-1\}$ be coprime to W. Then there exists a D-pseudorandom measure $\nu : \mathbb{Z}_{N'} \to \mathbb{R}^+$ (depending on $z_1,\dots,z_t$) which obeys the pointwise bounds*

$$1 + T^{z_1}\Lambda'_{b_1,W}(n) + \cdots + T^{z_t}\Lambda'_{b_t,W}(n) \ll_{D,C} \nu(n)$$

for all $n \in [N^{3/5}, N]$, where we identify n with an element of $\mathbb{Z}_{N'}$ in the obvious manner.

The proof of [17, Proposition 6.4] was presented in [17, Appendix D]. We now indicate the modifications necessary to that argument to obtain the more general Proposition 6.4'. The first modification we need to make is on page 1839, where we instead define the preliminary weight $\tilde{\nu} : [N] \to \mathbb{R}^+$ by setting

$$\tilde{\nu}(n) := \mathbb{E}_{i\in[t]}\frac{\phi(W)}{W}T^{z_i}\Lambda_{\chi,R,2}(Wn+b_i).$$

We have the bound

$$T^{z_i}\Lambda'_{b_i,W}(n) \ll_{C,D} \frac{\phi(W)}{W}T^{z_i}\Lambda_{\chi,R,2}(Wn+b_i) \tag{A.3}$$

for all $i \in [t]$ and all $n \in [N^{3/5}, N]$, analogous to that stated at the bottom of page 1839. The key observation here is that the left-hand side is only nonzero when $W(n+z_i)+b_i$ is a prime, in which case it equals $\frac{\phi(W)}{W}\log(W(n+z_i)+b_i) < \frac{2\phi(W)}{W}\log N$ (since $W \leqslant \log N, n \leqslant N$ and $z_i \leqslant N^{1.01}$). However if $n \in [N^{3/5}, N]$ then $W(n+z_i)+b_i \geqslant N^{3/5}$, and so if the sieve level γ used in the definition of $\Lambda_{\chi,R,2}$ satisfies $\gamma < \frac{3}{5}$ then the right-hand side is $\frac{\phi(W)}{W}\log R$. Since $R = N^\gamma$ and γ depends only on C, D (see halfway up page 1839), (A.3) follows.

As in [17, Appendix D], we then transfer to $\mathbb{Z}_{N'}$ by setting $\nu(n) := \frac{1}{2} + \frac{1}{2}\tilde{\nu}(n)$ when $n \in [N]$ and $\nu(n) := 1$ otherwise.

We then need to go back and modify the proof of [17, Theorem D.3] so that it applies with $T^{z_i}\Lambda_{\chi_i,R,a_i}$ replacing Λ_{χ_i,R,a_i}. Equivalently, we need to establish this theorem with only the weak bound $|\psi_i(0)| \ll N^{1.01}$ on the constant terms of the forms ψ_i, rather than the stronger bound $\|\Psi\|_N \leqslant L$ assumed there. In fact, no bound on the $\psi_i(0)$ is required in this part of the argument at all. The first place in that argument that the assumption $\|\Psi\|_N \leqslant L$ is used is in page 1833, where it is asserted that $\alpha(p,B) = \mathbb{E}_{\vec{n}\in\mathbb{Z}_p^d} 1_{p|\psi_i(\vec{n})}$

is equal to $1/p$ if $p \geqslant p_0(t, d, L)$ is sufficiently large. It is easy to see that the bound here depends only on the sizes of the coefficients in the homogeneous parts of ψ_i. The second place that this assumption is used is on page 1834, in the appeal to [17, Lemma 1.3]. As it happens only two of the three conclusions of this lemma as stated are valid under the weaker assumption: there is a superfluous statement about what happens for $p > C(d, t, L)N$ which fails in our present context, but which is not needed for the applications in [17, Appendix D]. An appropriately modified version of the lemma is the following.

Lemma 1.3'. *Suppose that $\Psi = (\psi_1, \dots, \psi_t)$ is a system of linear forms such that the homogeneous parts $\dot{\Psi} = (\dot{\psi}_1, \dots, \dot{\psi}_t)$ satisfy $\|\dot{\Psi}\|_N \leqslant L$. Then the local factors β_p satisfy $\beta_p = 1+O_{t,d,L}(p^{-1})$. If, furthermore, no two of the forms $\dot{\psi}_i$ are parallel then $\beta_p = 1 + O_{t,d,L}(p^{-2})$.*

This lemma, whose proof is the same as that of [17, Lemma 1.3], applies equally well on page 1834. The rest of the proof of [17, Theorem D.3] goes through unchanged.

The deduction of the linear forms conditions for $\tilde{\nu}$ now proceeds exactly as on pages 1840, with Λ_{χ_i,R,a_i} replaced by its shifted variant $T^{z_i}\Lambda_{\chi_i,R,a_i}$ whenever necessary.

The proof of the correlation conditions for $\tilde{\nu}$, starting at the bottom of page 1840, needs to be tweaked a little[8]. Instead of the bound at the bottom of page 1840, we must establish a variant with shifts, namely

$$\left(\frac{\phi(W)}{W}\right)^m \left(\sum_{n\in I}\prod_{j\in[m]} \Lambda_{\chi,R,2}(W(n+h_j) + b_{i_j} + Wz_{i_j})\right) \ll N \sum_{1\leqslant j<j'\leqslant m} \tau(h_j - h_{j'})$$

whenever $i_1, \dots, i_m \in [t]$. Here, the function τ is required to satisfy $\mathbb{E}_{n\in[-N,N]}\tau(n)^q \ll_q 1$. In the argument on page 1841, the set P_Ψ is now the set of primes dividing $W(h_j - h_j + z_{i_j} - z_{i'_j}) + b_{i_j} - b_{i_{j'}}$ for some $1 \leqslant j < j' \leqslant m$, and we define

$$\tau(n) := \sum_{1\leqslant j<j'\leqslant m} \exp\left(O(1) \sum_{\substack{p>w \\ p\mid Wn+W(z_{i_j}-z_{i_{j'}})+(b_{i_j}-b_{i_{j'}})}} \frac{1}{p^{1/2}}\right).$$

It now suffices to prove the bound

$$\mathbb{E}_{n\in[N]} \exp\left(q \sum_{\substack{p>w \\ p\mid Wn+h}} \frac{1}{p^{1/2}}\right) \ll_q 1$$

uniformly for all $h = O(N^{1.02})$. This is the same as the estimate at the bottom of page 1841, only there we had the stronger assumption $h = O(W)$. The only difference this makes to the argument is that the third displayed equation on page 1842 (which it is our task to prove) only comes with the weaker constraint $d = O(N^{1.02})$, that is to say we must show

$$\sum_{\substack{(d,W)=1 \\ d=O(N^{1.02})}} d^{-1/4} \sum_{\substack{n\in[N] \\ d\mid Wn+h}} 1 \ll N,$$

[8] Note, however, that by the work of Conlon, Fox and Zhao [4] one could in principle dispense with the need for this condition entirely.

whereas before we had $d = O(WN)$. However, the proof of this slightly stronger bound is the same: using the bound

$$\sum_{\substack{n\in[N] \\ d|Wn+h}} 1 \ll 1 + \frac{N}{d},$$

it reduces to

$$\sum_{d=O(N^{1.02})} d^{-1/4}\left(1 + \frac{N}{d}\right) \ll N,$$

a true statement. This at last completes the proof of Proposition 6.4'.

We now continue with the arguments of [17, Chapter 7]. Using Proposition 6.4' in place of [17, Proposition 6.4], we see that the proof of (A.1), and hence of Theorem 7, reduces to establishing the bound

$$\|T^z \Lambda'_{b,W} - 1\|_{U^{s+1}[N]} = o_{s,B}(1)$$

uniformly for all $b \in \{0, 1, \ldots, W-1\}$ and for all shifts z with $|z| \leqslant LN \log^B N$.

By the arguments of [17, Section 10] (but using Proposition 6.4' in place of [17, Proposition 6.4]) we can reduce to proving the bound

$$\mathbb{E}_{n\in[N]}(T^z\Lambda'_{b,W}(n) - 1)\psi(n) = o_{\psi,B}(1)$$

for any s-step nilsequence $\psi(n) = F(g^n x)$, where the $o_\psi(1)$ term may depend on the underlying nilmanifold G/Γ and the Lipschitz constant of F but not on the nilrotation g.

Chapter 11 of [17] requires no change, and the only changes required to Chapter 12 up to the bottom of page 1804 are to replace $\Lambda^\sharp$ and $\Lambda^\flat$ by their shifted variants $T^z\Lambda^\sharp$ and $T^z\Lambda^\flat$. This reduces matters to establishing the two estimates

$$\left\|\frac{\phi(W)}{W} T^z \Lambda^\sharp(Wn+b) - 1\right\|_{U^{s+1}[N]} = o_s(1) \tag{A.4}$$

(the shifted analogue of (12.5) in [17]) and

$$\mathbb{E}_{n\in[N]} \frac{\phi(W)}{W} T^z\Lambda^\flat(Wn+b)\psi(n) = o_{\psi,B}(1) \tag{A.5}$$

for all nilsequences ψ (the shifted analogue of (12.4) in [17].

The proof of the first of these, (A.4), proceeds exactly as in the proof of (12.5) of [17], which is given on page 1842–1843. The only change required is to use the variant of [17, Theorem D.3] with shifts, the validity of which was noted above. For this argument, we do not need any bound on z.

Finally we turn to the estimate (A.5). The analysis of page 1805 may be easily adapted, with the result that it is enough to prove that

$$\mathbb{E}_{n\in[N]} T^{zW}\Lambda^\flat(n)\psi(n) = o_{\psi,B}(1).$$

This, however, follows immediately from (12.10) of [17], which asserted the bound

$$\left|\sum_{n\in[N]} \Lambda^\flat(n)\psi(n)\right| \ll_{\psi,A} N\log^{-A} N$$

for any A. In particular, taking $A = B + 2$ (and noting that $W = o(\log N)$ and $z \leqslant LN \log^B N$) we have

$$\Big| \sum_{1 \leqslant n \leqslant N+zW} \Lambda^\flat(n)\psi(n) \Big| = o_{\psi,B}(N)$$

and

$$\Big| \sum_{1 \leqslant n \leqslant N} \Lambda^\flat(n)\psi(n) \Big| = o_{\psi,B}(N).$$

Subtracting these two estimates gives the result.

References

[1] A. Balog, *The prime k-tuplets conjecture on average*, Analytic number theory (Allerton Park, IL, 1989), 4775, Progr. Math., 85, Birkhäuser Boston, Boston, MA, 1990.

[2] R. C. Baker, G. Harman and J. Pintz, *The difference between consecutive primes. II.*, Proc. London Math. Soc. (3) **83** (2001), no. 3, 532–562.

[3] N. G. de Bruijn, *On the number of positive integers $\leqslant x$ and free of prime factors $> y$.* Nederl. Acad. Wetensch. Proc. Ser. A. **54** (1951) 50–60.

[4] D. Conlon, J. Fox and Y. Zhao, *A relative Szemerédi theorem,* Geom. Funct. Anal. **25** (2015), 733–762.

[5] H. Cramér, *Some theorems concerning prime numbers*, Ark. Mat. Astr. Fys. **15** (1920), 1–33.

[6] H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers,* Acta Arith. **2** (1936), 23–46.

[7] H. Davenport, *Multiplicative number theory*, 3rd ed., Graduate Texts in Mathematics vol. 74, Springer-Verlag, New York, 2000.

[8] L. E. Dickson, History of the theory of numbers, vol. III, Carnegie Inst. of Washington, Washington, DC 1919, 1920, 1923.

[9] P. Erdős, *On the difference of consecutive primes*, Quart. J. Math. Oxford Ser. **6** (1935), 124–128.

[10] P. Erdős, *Some of my favourite unsolved problems*, in A Tribute to Paul Erdős (A. Baker, B. Bollobás, A. Hajnal, eds.), Cambridge Univ. Press, 1990, pp. 467–478.

[11] K. Ford, B. Green, S. Konyagin, J. Maynard, T. Tao, *Long gaps between primes*, preprint.

[12] J. Friedlander, H. Iwaniec, Opera de cribro. American Mathematical Society Colloquium Publications, 57. American Mathematical Society, Providence, RI, 2010.

[13] W. T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, GAFA **8** (1998), 529–551.

[14] A. Granville, *Harald Cramér and the distribution of prime numbers*, Scandanavian Actuarial J. **1** (1995), 12–28.

[15] B. J. Green and T. C. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. **167** (2008), 481–547.

[16] B. J. Green and T. C. Tao, *The quantitative behaviour of polynomial orbits on nilmanifolds*, Annals of Math. **175** (2012), no. 2, 465–540.

[17] B. J. Green and T. C. Tao, *Linear equations in primes,* Annals of Math. **171** (2010), no. 3, 1753–1850.

[18] B. J. Green and T. C. Tao and T. Ziegler, *An inverse theorem for the Gowers U^4-norm,* Glasg. Math. J. **53** (2011), no. 1, 1–50.

[19] B. J. Green, T. C. Tao and T. Ziegler, *An inverse theorem for the Gowers $U^{s+1}[N]$-norm*, Annals of Math. **176** (2012), 1231–1372.

[20] H. Halberstam and H.-E. Richert, *Sieve Methods*, Academic Press, London, 1974.

[21] D. R. Heath-Brown, *Gaps between primes, and the pair correlation of zeros of the zeta function*, Acta Arith. **41** (1982), no. 1, 85–99.

[22] G. H. Hardy, J. E. Littlewood, *Some Problems of 'Partitio Numerorum.' III. On the Expression of a Number as a Sum of Primes*, Acta Math. **44** (1923), 1–70.

[23] H. Iwaniec, *On the problem of Jacobsthal*, Demonstratio Math. **11** (1978), 225–231.

[24] H. Maier and C. Pomerance, *Unusually large gaps between consecutive primes*. Trans. Amer. Math. Soc. **322** (1990), no. 1, 201–237.

[25] J. Maynard, *Small gaps between primes*, Ann. of Math. (2) **181** (2015), no. 1, 383–413.

[26] J. Maynard, *Large gaps between primes*, preprint.

[27] J. Pintz, *Very large gaps between consecutive primes.* J. Number Theory **63** (1997), no. 2, 286–301.

[28] R. A. Rankin, *The difference between consecutive prime numbers*, J. London Math. Soc. **13** (1938), 242–247.

[29] R. A. Rankin, *The difference between consecutive prime numbers. V,* Proc. Edinburgh Math. Soc. (2) 13 (1962/63), 331–332.

[30] A. Schönhage, *Eine Bemerkung zur Konstruktion grosser Primzahllücken*, Arch. Math. **14** (1963), 29–30.

[31] T. Oliveira e Silva, S. Herzog, S. Pardi, *Empirical verification of the even Goldbach conjecture and computation of prime gaps up to* 4×10^{18}, Math. Comp. **83** (2014), 2033–2060.

[32] B. Szegedy, *Gowers norms, regularization and limits of functions on abelian groups*, preprint.

[33] E. Westzynthius, *Über die Verteilung der Zahlen, die zu den* n *ersten Primzahlen teilerfremd sind,* Commentationes Physico–Mathematicae, Societas Scientarium Fennica, Helsingfors **5**, no. 25, (1931) 1–37.

DEPARTMENT OF MATHEMATICS, 1409 WEST GREEN STREET, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, URBANA, IL 61801, USA
E-mail address: ford@math.uiuc.edu

MATHEMATICAL INSTITUTE, RADCLIFFE OBSERVATORY QUARTER, WOODSTOCK ROAD, OXFORD OX2 6GG, ENGLAND
E-mail address: ben.green@maths.ox.ac.uk

STEKLOV MATHEMATICAL INSTITUTE, 8 GUBKIN STREET, MOSCOW, 119991, RUSSIA
E-mail address: konyagin@mi.ras.ru

DEPARTMENT OF MATHEMATICS, UCLA, 405 HILGARD AVE, LOS ANGELES CA 90095, USA
E-mail address: tao@math.ucla.edu