.

LONG GAPS IN SIEVED SETS

KEVIN FORD, SERGEI KONYAGIN, JAMES MAYNARD, CARL POMERANCE, AND TERENCE TAO

ABSTRACT. For each prime p, let $I_p \subset \mathbb{Z}/p\mathbb{Z}$ denote a collection of residue classes modulo p such that the cardinalities $|I_p|$ are bounded and about 1 on average. We show that for sufficiently large x, the sifted set $\{n \in \mathbb{Z} : n \pmod p \notin I_p \text{ for all } p \leqslant x\}$ contains gaps of size at least $x(\log x)^\delta$ where $\delta > 0$ depends only on the density of primes for which $I_p \neq \emptyset$. This improves on the "trivial" bound of $\gg x$. As a consequence, for any non-constant polynomial $f : \mathbb{Z} \to \mathbb{Z}$ with positive leading coefficient, the set $\{n \leqslant X : f(n) \text{ composite}\}$ contains an interval of consecutive integers of length $\geqslant (\log X)(\log\log X)^\delta$ for sufficiently large X, where $\delta > 0$ depends only on the degree of f.

1. INTRODUCTION

It is well-known that the sieve of Eratosthenes sometimes removes unusually long strings of consecutive integers, and this implies that the sequence of primes occasionally has much longer gaps than the average spacing. It might be expected that similar methods would show analogous results for other sets undergoing a sieve, such as sets defined by polynomials. For example, we know that the number of $n \leqslant x$ with $n^2 + 1$ prime is $O(x/\log x)$, so an immediate corollary is that there are intervals of length $\gg \log x$ below x where n^2+1 is composite for each n in the interval. Can we do better? A simple averaging argument is not useful, since the $O(x/\log x)$ bound for the count is conjecturally best possible. In addition, there unfortunately appear to be fundamental obstructions to adapting the methods used to locate large gaps in the Eratosthenes sieve to this situation.

In this paper we introduce a new method which substantially improves upon the trivial bound for these polynomial sets, and applies to more general sieving situations. We consider the set of integers remaining after applying a "one-dimensional" sieve, and show that this sieved set contains some unusually large gaps. To state our theorem precisely we require the following definition. The symbol p always denotes a prime.

Definition 1 (Sieving System)**.** *A* sieving system *is a collection $\mathcal{I}$ of sets $I_p \subset \mathbb{Z}/p\mathbb{Z}$ of residue classes modulo p for each prime p. Moreover, we have the following definitions.*

- *(Non-degeneracy) We say that the sieving system is* non-degenerate *if $|I_p| \leqslant p - 1$ for all p.*
- *(B-Boundedness) Given $B > 0$, we say that the sieving system is B-*bounded *if*

$$|I_p| \leqslant B \text{ for all primes } p. \tag{1.1}$$

Date: January 3, 2020.

KF was supported by National Science Foundation grant DMS-1501982. JM was supported by a Clay Research Fellowship and a Fellowship of Magdalen College, Oxford. 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. Part of this work was carried out at MSRI, Berkeley during the Spring semester of 2017, supported in part by NSF grant DMS-1440140. We thank the anonymous referees for many useful suggestions.

2010 Mathematics Subject Classification: Primary 11N35, 11N32, 11B05.

Keywords and phrases: gaps, prime values of polynomials, sieves.

- *(One-dimensionality) We say that the sieving system is* one-dimensional *if we have the weighted Mertens-type product estimate*

$$\prod_{p\leqslant x}\left(1-\frac{|I_p|}{p}\right)\sim\frac{C_1}{\log x}\qquad(x\to\infty), \tag{1.2}$$

for some constant $C_1>0$.

- *(ρ-supportedness) Given $\rho>0$, we say that the sieving system system is ρ-*supported *if the density of primes with $|I_p|\geqslant 1$ equals ρ, that is,*

$$\lim_{x\to\infty}\frac{|\{p\leqslant x:|I_p|\geqslant 1\}|}{x/\log x}=\rho. \tag{1.3}$$

Roughly speaking, a "sieving system" which is non-degenerate, B-bounded, 1-dimensional and ρ-supported specifies certain residue classes for each prime p, such that there is roughly 1 residue class per prime on average, and if we remove all integers in these residue classes the resulting set is not too erratic.

Given such a sieving system $\mathcal{I}$, our main object of study is the *sifted set*

$$S_x=S_x(\mathcal{I}):=\mathbb{Z}\setminus\bigcup_{p\leqslant x}I_p,$$

of integers which are not contained in any of the residue classes specified by the I_p for $p\leqslant x$. If $|I_p|=p$ for some $p\leqslant x$ (the degenerate case), then clearly S_x is empty. Otherwise, S_x is a $P(x)$-periodic set with density $\sigma(x)$, where $P(x)$ and $\sigma(x)$ are defined as

$$P(x):=\prod_{\substack{p\leqslant x\\ I_p\neq\emptyset}}p,\qquad \sigma(x):=\prod_{p\leqslant x}\left(1-\frac{|I_p|}{p}\right).$$

We also note that $S_x\supseteq S_y$ if $x\leqslant y$. With this set-up we can now state our main theorem.

Theorem 1 (Main theorem)**.** *Let $\mathcal{I}$ be a non-degenerate, B-bounded, one-dimensional, ρ-supported sieving system with $\rho>0$. Define*

$$C(\rho):=\sup\Big\{\delta\in(0,1/2):\ \frac{(4+\delta)\cdot 10^{2\delta}}{\log(1/(2\delta))}<\rho\Big\}. \tag{1.4}$$

The sifted set S_x contains a gap of length at least $x(\log x)^{C(\rho)-o(1)}$, where the rate of decay of the $o(1)$ bound depends on $\mathcal{I}$. Moreover, $C(\rho)>e^{-1-4/\rho}$.

Remark 1. We note that since $\mathcal{I}$ is one-dimensional, we must have that

$$\rho\geqslant\frac{1}{B}.$$

(So, for example, the positivity of ρ follows from the property that $\mathcal{I}$ is B-bounded.) The value of C_1 in (1.2), which has no importance for our arguments, depends on the behavior of $|I_p|$ for small p, and can have great variation.

Condition (1.3) is used primarily to construct large sets of primes with $I_p\neq\emptyset$ in very short intervals, see (2.8) below. It is possible to weaken (1.3) further, e.g. so that (2.8) holds for most scales H instead of all H, however this would further complicate our argument. All of the canonical examples satisfy (1.3).

There is a straightforward argument that shows that S_x must have gaps of length $\gg x$, for x sufficiently large in terms of $\mathcal{I}$ — see Remark 5 below. Theorem 1 improves over this bound by a positive power of $\log x$, and it is the fact that we get a non-trivial result in this level of generality which is the main point of the Theorem. It is likely that with more effort one could improve the bounds on the constant $C(\rho)$; our main interest is that this is an explicit positive constant depending only on ρ. We now demonstrate applications of the theorem via several examples.

Example 1 (Gaps between primes). The "Eratosthenes" sieving system is the system with $I_p = \{0\}$ for all p, and it is non-degenerate, 1-bounded, one-dimensional and 1-supported. We have

$$\{\sqrt{X} < p \leqslant X : p \text{ prime}\} = S_{\sqrt{X}} \cap (\sqrt{X}, X]. \tag{1.5}$$

Since $S_x \supseteq S_{\sqrt{X}}$ if $x \leqslant \sqrt{X}$, any large gap in S_x implies a large gap in $S_{\sqrt{X}}$. Since S_x is $P(x)$-periodic, if it contains a large gap then it must contain one in the interval $[\sqrt{X}, X]$ if $P(x) \leqslant X - \sqrt{X}$. Thus, choosing $x \approx \log X$ maximally such that $P(x) \leqslant X - \sqrt{X}$, we see that Theorem 1 implies that there is a prime gap in $[\sqrt{X}, X]$ of size

$$\gg (\log X)(\log\log X)^{C(1)-o(1)} \gg (\log X)(\log\log X)^{1/128},$$

on numerically calculating that $C(1) > 1/128$ (the limit of our type of method appears to be an exponent $1/e$; see Remark 9 in Section 4). This is stronger than the trivial bound of $(1+o(1))\log X$, which is immediate from the Prime Number Theorem, but is worse than the current best bounds for this problem. Indeed, the problem of finding large gaps between consecutive primes has a long history, and it is currently known that gaps of size

$$\gg \log X \, \frac{\log\log X \, \log\log\log\log X}{\log\log\log X} \tag{1.6}$$

exist below X if X is large enough, a recent result of Ford, Green, Konyagin, Maynard, and Tao [5]. The key interest is that Theorem 1 applies to much more general sieving situations, to which it appears difficult to adapt the previous techniques, and gives a different method of proof to these previous results. We will discuss the reasons for this in detail below.

Example 2 (Gaps between prime values of polynomials). Given a polynomial $f : \mathbb{Z} \to \mathbb{Z}$ of degree $d \geqslant 1$, consider the system $\mathcal{I}$ with $I_p = \emptyset$ for $p \leqslant d$ and

$$I_p := \{n \in \mathbb{Z}/p\mathbb{Z} : f(n) \equiv 0 \pmod p\}$$

for $p > d$. The polynomial need not have integer coefficients, e.g. $f(n) = \frac{n^7-n+7}{7}$ satisfies the hypotheses of Theorem 1. By Pólya's theorem [11], f is integer valued at integers if and only if f has the form $f(x) = \sum_{j=0}^{d} a_j \binom{x}{j}$ with every $a_j \in \mathbb{Z}$. In particular, $d!f(y) \in \mathbb{Z}[y]$ and thus the sieving system is well-defined.

By Lagrange's theorem, $|I_p| \leqslant d < p$ for all $p > d$, and hence the system is non-degenerate and d-bounded. For irreducible f, the one-dimensionality (1.2) with strong error term follow quickly from Landau's Prime Ideal Theorem [10] (see also [4, pp. 35–36]), while (1.3), the ρ-supportedness of the system with $\rho \geqslant 1/d$, follows from the Chebotarev Density Theorem [3] (see also [9]). As a variant of (1.5), we observe that

$$\{n \in \mathbb{N} : f(n) > x, f(n) \text{ prime}\} \subset S_x$$

for any $x > 1$. Now set $x := \frac{1}{2}\log X$. By Theorem 1, the set S_x contains a gap of length $\gg (\log X)(\log\log X)^{C(1/d)-o(1)}$. The period of this set, $P(x)$, is $X^{1/2+o(1)}$ by the Prime Number Theorem. Thus, this set contains such a long gap inside the interval $[X/2, X]$. Assuming that f has a positive leading coefficient and that X is large, on the interval $[X/2, X]$ we have $f(n) > x$, and so $f(n)$ is composite for every $n \in [X/2, X]\backslash S_x$. We thus obtain the following.

Corollary 1. *Let $f : \mathbb{Z} \to \mathbb{Z}$ be a polynomial of degree $d \geqslant 1$ with positive leading term. Then for sufficiently large X, there is a string of consecutive natural numbers $n \in [1, X]$ of length $\geqslant (\log X)(\log\log X)^{C(1/d)-o(1)}$ for which $f(n)$ is composite, where $C(1/d) > e^{-(4d+1)}$ is the constant of Theorem 1.*

Note that Corollary 1 includes the trivial "degenerate" cases, when either f is reducible, or there is some prime p with $|I_p| = p$, since then essentially all values of f are composite.

When f is irreducible, has degree two or greater, and the sieving system corresponding to f is non-degenerate, it is still an open conjecture (of Bunyakovsky [2]) that there are infinitely many integers n for which $f(n)$ is prime. Moreover it is believed (see the conjecture of Bateman and Horn [1]) that the density of these prime values on $[X/2, X]$ is $\asymp_f 1/\log X$, and so the gaps of Corollary 1 would be unusually large compared to the average gap of size $\asymp_f \log X$. We do not address these conjectures at all in this paper. Of course, in the unlikely event that Bunyakovsky's conjecture was false and there were only finitely many prime values of f, Corollary 1 would be worse than the trivial bound.

Remark 2. Let G be the Galois group of f, realized canonically as a subgroup of the symmetric group $\mathfrak{S}_d$. By the Chebotarev Density Theorem [3] (see also [9]), we may take ρ equal to the proportion of elements of G with at least one fixed point, which lies in $[\frac{1}{d}, 1)$. We have $\rho = 1/d$ for many polynomials, e.g. $x^{2^k} + 1$, but ρ is much larger generically. It is known since van der Waerden [13] that a random irreducible polynomial of degree d will have Galois group $\mathfrak{S}_d$ with high probability[1]. In this case ρ is the proportion of elements of $\mathfrak{S}_d$ with a fixed point. This is the classical derangement problem, and we have for such polynomials

$$\rho = \rho_d := \sum_{k=1}^{d} \frac{(-1)^{k+1}}{k!}.$$

In particular, $\rho_d \geqslant 1/2$, $\rho_d \geqslant \frac{5}{8}$ for $d \geqslant 3$ and $\lim_{d\to\infty} \rho_d = 1 - 1/e$. A calculation reveals that

$$C(1/2) > \frac{1}{6001}. \tag{1.7}$$

Since $C(\rho)$ is increasing with ρ, we thus have the following corollary.

Corollary 2. *Let $f : \mathbb{Z} \to \mathbb{Z}$ be a polynomial of degree $d \geqslant 2$ with positive leading term, irreducible over $\mathbb{Q}$, and with full Galois group $\mathfrak{S}_d$. Then for all sufficiently large X, there is a string of consecutive natural numbers $n \in [1, X]$ of length $\geqslant \log X(\log\log X)^{1/6001}$ for which $f(n)$ is composite.*

[1]Specifically, fix the degree d and let the coefficients of f be chosen randomly and uniformly from $[-N, N] \cap \mathbb{Z}$. Then, as $N \to \infty$, the probability that f is irreducible and has Galois group $\mathfrak{S}_d$ tends to 1.

Example 3. A simple example to keep in mind is $f(n) = n^2 + 1$. In this case, $I_2 = \{1\}$, $I_p = \emptyset$ is empty for $p \equiv 3 \pmod 4$, and $I_p = \{\iota_p, -\iota_p\}$ for $p \equiv 1 \pmod 4$, where $\iota_p \in \mathbb{Z}/p\mathbb{Z}$ is one of the square roots of -1. Here one can use the Prime Number Theorem in arithmetic progressions rather than the Prime Ideal theorem to establish one-dimensionality and the ρ-supportedness with $\rho = 1/2$. For this example (and for any quadratic polynomial), Theorem 1 implies the existence of consecutive composite strings of length $\gg (\log X)(\log\log X)^{C(1/2)-o(1)} \gg (\log X)(\log\log X)^{1/6001}$ (using (1.7) again). It is certain that further numerical improvements are possible.

Theorem 1 has another application, to a problem on the coprimality of consecutive values of polynomials.

Corollary 3. *Let $f : \mathbb{Z} \to \mathbb{Z}$ be a non-constant polynomial. Then there exists an integer $G_f \geqslant 2$ such that for any integer $k \geqslant G_f$ there are infinitely many integers $n \geqslant 0$ with the property that none of the numbers $f(n+1), \dots, f(n+k)$ are coprime to all the others.*

For linear polynomials the result of the corollary is well-known, and not difficult to prove; for quadratic and cubic polynomials in $\mathbb{Z}[x]$, the result was only proven recently by Sanna and Szikszai [12]. The remaining cases of polynomials of degree four and higher appears to be new.

Proof. Let $d = \deg f$. Then $d!f(y) \in \mathbb{Z}[y]$. Let $f_0(y) \in \mathbb{Z}[y]$ be a primitive irreducible factor of $d!f(y)$. If $p > d$ is a prime and $p \mid f_0(m)$ for some integer m, then $p \mid f(m)$. So it will suffice to consider the case that f is irreducible and show in this case that for all large k there are infinitely many $n \geqslant 0$ such that for each $i \in \{1, \dots, k\}$ there is some $j \in \{1, \dots, k\}$ with $j \neq i$ and $\gcd(f(n+i), f(n+j))$ divisible by some prime $> d$.

Again, we consider the system $\mathcal{I}$ defined by $I_p = \emptyset$ for $p \leqslant d$ and for $p > d$ we take

$$I_p := \{n \in \mathbb{Z}/p\mathbb{Z} : f(n) \equiv 0 \pmod p\}.$$

By Theorem 1, for all large numbers x the set S_x contains a gap of length $\geqslant k = \lfloor 2x \rfloor$. Thus, there are infinitely many n such that each $f(n+1), \dots, f(n+k)$ has a prime factor p with $d < p \leqslant x$. For each $i \in \{1, \dots, k\}$, take a prime factor p of $f(n+i)$ with $d < p \leqslant x$. Since $k = \lfloor 2x \rfloor$, $p \leqslant x$ and $I_p \neq \emptyset$, it must be that p divides at least two terms of the sequence $f(n+1), \dots, f(n+k)$, thus proving the assertion. □

Remark 3. Our proof of Corollary 3 above requires only a very weak version of Theorem 1. It is not clear, however, that a trivial argument of the type presented below in Remark 5 can yield a gap of size at least $2x$ when the degree of f is large.

Remark 4. The conclusion of Theorem 1 is equivalent to the existence, for any $\delta < C(\rho)$, of some $b \in \mathbb{Z}/P(x)\mathbb{Z}$ with

$$(S_x + b) \cap [1, x(\log x)^\delta] = \emptyset,$$

provided x is sufficiently large in terms of δ. Here $S_x + b := \{s + b : s \in S_x\}$.

Remark 5. The conclusion of Theorem 1 should be compared with the "trivial" bound: there is a constant $c' > 0$ such that for each sufficiently large x, there is some integer b with

$$(S_x + b) \cap [1, c'x] = \emptyset. \tag{1.8}$$

We now sketch the proof of (1.8). Firstly, we see that we may assume that x is large. Then by (1.2) it follows that there is some b modulo $P(x/2)$ for which $\mathcal{A} := (S_{x/2} + b) \cap [1, \frac{\rho x}{8C_1}]$ satisfies

$|\mathcal{A}| \leqslant \frac{\rho x}{4 \log x}$. On the other hand, by (1.3) for any fixed $\varepsilon > 0$ we have

$$\#\{x/2 < q \leqslant x : |I_q| \geqslant 1\} \geqslant \left(\frac{\rho}{2} - \varepsilon\right) \frac{x}{\log x} \tag{1.9}$$

for large x. Hence, we may perform a "clean up stage" in which we pair up each element $a \in \mathcal{A}$ with a unique prime $q = q_a \in (x/2, x]$ for which $|I_q| \geqslant 1$. For each such pair a, q_a let $v_a \in I_{q_a}$ and suppose that $b \equiv a - v_a \pmod q$. It follows that $(S_x + b) \cap [1, \frac{\rho x}{8C_1}] = \emptyset$, proving (1.8).

Remark 6. The hypothesis (1.1) is an important assumption in our treatment of certain error terms; see Lemma 5.1 below. It is possible to relax this hypothesis with more sophisticated arguments, and several steps of the argument could be established with slightly weaker assumptions.

The formula (1.2) say that $|I_p|$ has average 1 in a weak sense, and is similar to the usual condition defining a *one-dimensional* sieve (see e.g. [6, Sections 5.5, 6.7]). Most of our arguments have counterparts if the one-dimensional hypothesis (1.2) is replaced by another dimension, but in those cases the bounds we could obtain were inferior to what could be obtained by the "trivial" argument; see for instance Remark 7 below.

1.1. **Comparisons of methods.** Recall from Example 1 that for the Eratosthenes sieving system $I_p = \{0\}$, previous methods were able to deduce stronger variants of Theorem 1. We now explain why these methods appear difficult to adapt to more general sieving systems.

In the Eratosthenes sieving system it is clear that S_x avoids the interval $[2, x]$, which already gives the "trivial" lower bound $j(P(x)) \geqslant x - 2$. All of the improvements to this bound in previous literature (including those in [5]) rely on a variant of the following observation: if $x \geqslant z \geqslant 2$, then the sifted set

$$S_{z,x} = \mathbb{N} \setminus \bigcup_{z < p \leqslant x} I_p, \tag{1.10}$$

when restricted to the interval $[1, y)$ with y slightly larger than x, only consists of numbers of the form a or ap, where p is a prime in $(x, y]$, and a is *z-smooth* (or *z-friable*), which means that no prime factor of a exceeds z. Moreover, z-smooth numbers are much rarer than one would expect from naive sieving heuristics (if z is suitably small), but numbers of the form ap must have a less than y/x, which is also a rare factorization (if y is only slightly larger than x). Thus the number of elements of $S_{z,x}$ in $[1, y)$ is unusually small. It is the fact that we can identify this interval containing unusually few integers after sieving by the "medium-sized" primes which is the key ingredient allowing one to improve on the trivial bound.

The most recent works on this problem then try to show as efficiently as possible that one choose b (a multiple of $\prod_{z<p\leqslant x} p$) such that $(b + S_{2x}) \cap [1, y) = \emptyset$, and so we can sieve out out these few remaining elements of $[1, y)$. This then implies the existence of a large gap of size y in S_{2x}. However, if we did not already know that there were few elements in $[1, y)$, then these methods would not produce a non-trivial bound.

Unfortunately, when considering the more general sieving systems of Definition 1 in which the cardinalities $|I_p|$ are allowed to vanish for many primes p, bounds for smooth numbers cannot be used to show that $S_{z,x}$ contains an interval with unusually few elements. Without this crucial step the existing methods only yield the trivial lower bound of $\gg x$ for the gap size. Moreover, for a general sieving system which is ρ-supported with $\rho < 1$, we expect that *no* such reasonably long

interval containing so few elements will exist in $S_{z,x}$, meaning that this feature is genuinely unique to the Eratosthenes sieving system.

We overcome this obstacle by using a rather different method. Rather than attempting to do unusually well with the medium sized primes $p < x/(\log x)^{1/2}$, we instead will make random choices, and only obtain results comparable to the trivial bound. We obtain an improvement over the trivial bound by working harder with the larger primes $p \in [x/(\log x)^{1/2}, x]$, showing that for each of these larger primes we can actually remove more elements that one would typically expect by choosing the residue class carefully. In order to make sure these choices do not interfere with each other too much, we make the choices randomly in several stages, where the random choice is conditional on the previous stages.

The basic idea is similar to how recent papers (e.g. [5]) have exploited the large primes to sieve efficiently. In those papers one needed estimates of tuples of linear forms taking many prime values frequently, here we just need to show the existence of suitable residue classes containing unusually many unsieved integers. However, in the new set-up we require rather stronger quantitative bounds than is available for tuples of prime values - our method would completely fail to improve over the trivial bound if we were not able to obtain close-to-optimal quantitative results. This strategy is discussed in more detail in the next section.

Remark 7. Unfortunately our methods only seem to give good results in the one-dimensional case. Consider for instance the set $\{n \in \mathcal{P} : n + 2 \in \mathcal{P}\}$ of (the lower) twin primes. This corresponds to a two-dimensional system in which $I_p = \{0 \pmod p, 2 \pmod p\}$ for all primes p. The "trivial" bound coming from these methods would give a bound of $\gg \log X \log\log X$ for the largest gap between lower twin primes up to X (or between the largest such twin prime and X), and one could possibly hope to improve this bound by a small power of $\log\log X$ using a variant of the methods in this paper. However, a sieve upper bound (e.g., [7, Cor. 2.4.1]) combined with the pigeonhole principle already gives a bound of $\gg \log^2 X$ in this case.

1.2. **Notation.** From now on, we shall fix a non-degenerate, B-bounded, one-dimensional, ρ-supported sieving system $\mathcal{I}$.

We use $X \ll Y, Y \gg X$, or $X = O(Y)$ to denote the estimate $|X| \leqslant CY$ for some constant $C > 0$, and write $X \asymp Y$ for $X \ll Y \ll X$. Throughout the remainder of the paper, all implied constants in $O(\cdot)$ and related order estimates may depend on $\mathcal{I}$, in particular on the constants B, ρ, C_1. Moreover, implied constants will also be allowed to depend on quantities δ, M, K, and ξ which we specify in the next section. We also assume that the quantity x is sufficiently large in terms of all of these parameters.

The notation $X = o(Y)$ as $x \to \infty$ means $\lim_{x\to\infty} X/Y = 0$ (holding other parameters fixed).

If S is a statement, we use 1_S to denote its indicator, thus $1_S = 1$ when S is true and $1_S = 0$ when S is false.

We will rely on probabilistic methods in this paper. Boldface symbols such as $\mathbf{n}$, $\mathbf{S}$, $\boldsymbol{\lambda}$, etc. denote random variables (which may be real numbers, random sets, random functions, etc.). Most of these random variables will be discrete (in fact they will only take on finitely many values), so that we may ignore any technical issues of measurability; however it will be convenient to use some continuous random variables in the appendix. We use $\mathbb{P}(E)$ to denote the probability of a random event E, and $\mathbb{E}\mathbf{X}$ to denote the expectation of the random (real-valued) variable $\mathbf{X}$.

Unless specified, all sums are over the natural numbers. An exception is made for sums over the variables p or q (as well as variants such as p_1, p_2, etc.), which will always denote primes.

2. OUTLINE

In this section we describe the high-level strategy of proof, and perform two initial reductions on the problem, ultimately leaving one with the task of proving Theorem 2 below. Recall the definition (1.10) of the sifted set $S_{z,x}$ and define related quantities

$$P(z,x) := \prod_{\substack{z<p\leqslant x\\ |I_p|\geqslant 1}} p, \qquad \sigma(z,x) := \prod_{z<p\leqslant x}\left(1-\frac{|I_p|}{p}\right).$$

Suppose x is large (think of $x\to\infty$), and define

$$y := \lceil x(\log x)^{\delta}\rceil \tag{2.1}$$

and

$$z := \frac{y\log\log x}{(\log x)^{1/2}}, \tag{2.2}$$

where $\delta\in(0,1/2)$ satisfies $\delta<C(\rho)$. We recall from (1.4) that this is equivalent to

$$\frac{(4+\delta)\cdot 10^{2\delta}}{\log(1/(2\delta))}<\rho, \tag{2.3}$$

which is a condition that will arise naturally in the proof. Our goal is to show that $(S_x+b)\cap[1,y]=\emptyset$ for some b and to accomplish this with maximal δ such that (2.3) holds. For a general ρ, it is easy to see that

$$C(\rho)>e^{-1-4/\rho} \qquad (0<\rho\leqslant 1),$$

establishing the final claim in Theorem 1. Incidentally, $C(\rho)=\frac{1}{2}e^{-(4+o(1))/\rho}$ as $\rho\to 0^+$.

In the course of the proof, we will introduce three additional parameters: M is a fixed number slightly larger than 4, ξ is a real number slightly large than 1, and K is a very large integer; we will eventually take $\xi\to 1^+$ and $K\to\infty$. We adopt the convention that constants implied by $O(\cdot)$ and $\ll$ bounds may depend on δ, M, K, ξ, in addition to the parameters defining $\mathcal{I}$, that is ρ, B, C_1. Dependence on any other parameter will be stated explicitly.

We observe that a linear shift of any single set I_p (that is, replacing I_p by $c+I_p$ for some integer c) does not affect the structure of S_x. Thus, the same is true for linear shifts (depending on p) for any finite set of primes p. In particular, we may shift the sets I_p so that all nonempty sets I_p contain the zero element, without changing the structure of S_x. Therefore, we may assume without loss of generality that $0\in I_p$ whenever I_p is nonempty. By the Chinese Remainder Theorem, we may select b by choosing residue classes for b modulo primes $p\leqslant x$.

2.1. **Basic Strategy.**

For x large enough we have

$$1\leqslant z\leqslant x/2\leqslant x\leqslant y.$$

We will select the parameter b modulo the primes $p\leqslant x$ in three stages:

(1) (Uniform random stage) First, we choose b modulo $P(z)$ uniformly at random; equivalently, for each prime $p\leqslant z$ with $|I_p|\geqslant 1$, we choose $b \bmod p$ randomly with uniform probability, independently for each p.

(2) (Greedy stage) Secondly, choose b modulo $P(z, x/2)$ randomly, but dependent on the choice of b modulo $P(z)$. A bit more precisely, for each prime $q \in (z, x/2]$ with $|I_q| \geqslant 1$, we will select $b \equiv b_q \pmod q$ so that $\{b_q + kq : k \in \mathbb{Z}\} \cap [1, y]$ knocks out nearly as many elements of the random set $(S_z + b) \cap [1, y]$ as possible. Note that we are focusing only on those residues sifted by the element $0 \in I_q$, and ignoring all other possible elements of I_q. This simplifies our analysis considerably, but has the effect of making $C(\rho)$ decay rapidly as $\rho \to 0$.

(3) (Clean up stage) Thirdly, we choose b modulo primes $q \in (x/2, x]$ to ensure that the remaining elements $m \in (S_{x/2} + b) \cap [1, y]$ do not lie in $(S_x + b) \cap [1, y]$ by matching a unique prime $q = q(m)$ with $|I_q| \geqslant 1$ to each element m and setting $b \equiv m \pmod q$. (Again we use the single element $0 \in I_q$. Such a clean up stage is standard in this subject, for instance it was already used in the proof of (1.8).)

We then wish to show that there is a positive probability that the above random sieving procedure has $(S_x + b) \cap [1, y] = \emptyset$, which then clearly implies that there is a choice of b such that this is the case, giving Theorem 1. It is the second sieving stage above which is the key new content of this paper.

Following the argument used to show (1.8), and using (1.9), we can successfully show that there exists a b' such that $(S_x + b') \cap [1, y] = \emptyset$ after Stage (3) provided that we have suitably few elements after Stage (2). By (1.9) (a consequence of our hypothesis (1.3)), it is sufficient to show that there is a b such that

$$|(S_{x/2} + b) \cap [1, y]| \leqslant \Big(\frac{\rho}{2} - \varepsilon\Big) \frac{x}{\log x}. \tag{2.4}$$

After Stage (1), from (1.2) we see that the expected size of $|(S_z + b) \cap [1, y]|$ is $\sim \sigma(z)y \asymp \frac{y}{\log z} \sim \frac{y}{\log x}$. A random, uniform choice of b modulo primes $q \in (z, x/2]$ would only reduce the residual set by a factor $\prod_{z<p\leqslant x/2}(1 - |I_p|/p) \sim 1$ and would lead to a version of Theorem 1 with a gap of size $\asymp x$. Instead, we use a greedy algorithm to select $b \equiv b_q \pmod q$. By (2.1) and (2.2), the set $(b_q \bmod q) \cap [1, y]$ has size about $H := y/q$, with $(\log x)^\delta \ll H \ll (\log x)^{1/2}/\log\log x$. By considering the initial portion $(S_{H^M} + b)$ (for some fixed $M > 1$) of the sieving process, one can see (e.g. using the large sieve [6, Lemma 7.5 and Cor. 9.9] or Selberg's sieve [8, Sec. 1.2]) that the size of the intersection $(b_q \bmod q) \cap (S_{H^M} + b) \cap [1, y]$ must be somewhat smaller, namely of size

$$\ll \sigma(H)H \asymp \frac{H}{\log H}$$

by (1.2). We will show that there are choices for the residues b_q so that no further size reduction occurs when one sieves up to z instead of H^M, namely that

$$(S_{H^M} + b) \cap (b_q \bmod q) \cap [1, y] = (S_z + b) \cap (b_q \bmod q) \cap [1, y]. \tag{2.5}$$

Heuristically, each individual choice of b_q is expected to obey (2.5) with probability roughly

$$\sigma(H^M, z)^{H\sigma(H)},$$

but with our choice of parameters and (1.2), this quantity is substantially larger than $1/q$, and so there should be many possibilities for b_q for each q. By contrast, for most choices of b_q, the ratio of the left and right sides of (2.5) is about $\sigma(H^M, z) = \prod_{H^M<p\leqslant z}(1 - |I_p|/p) \sim \frac{\log H^M}{\log z}$, which is very small.

Remark 8. A simple way to perform the greedy stage would be to choose the b_q independently from one another for each q, conditional only on the first stage. One would then expect that that we will achieve (2.4) if $y = x(\log\log x)^{\rho-\varepsilon}$ instead of (2.1). This would give a non-trivial result which is weaker than Theorem 1. Indeed, imagine we had instead defined $z := x/J$ and $y := Lx$, where J and L lie in $\big[100, (\log x)^{1/3}\big]$. After Stage (1), we are left with a set $\mathcal{R}$ of approximately $y/\log x = Lx/\log x$ integers. The goal is to choose $b = b_q$ for primes $q \in (z, x/2]$ with nonempty I_q so that $b \bmod q$ knocks out $\approx (y/q)/(\log(y/q))$ elements of $\mathcal{R}$. For this to be possible, we must have $\sigma(H^M, z)^{H\sigma(H)} \geqslant 1/q$ for all $H \leqslant y/z = JL$, but this is true on account of $JL \leqslant (\log x)^{2/3}$. Assuming independence of all these steps (that is, for different q), the residual set after the greedy sieving has size

$$\lesssim |\mathcal{R}| \prod_{\substack{x/J<q\leqslant x/2\\ I_q\neq\emptyset}} \left(1 - \frac{(y/q)/\log(y/q)}{|\mathcal{R}|}\right) \approx \frac{Lx}{\log x} \prod_{\substack{x/J<q\leqslant x/2\\ I_q\neq\emptyset}} \left(1 - \frac{\log x}{q\log(y/q)}\right).$$

By the Prime Number Theorem and (1.3),

$$\sum_{\substack{x/J<q\leqslant x/2\\ I_q\neq\emptyset}} \frac{\log x}{q\log(y/q)} = \rho\int_{x/J}^{x/2} \frac{dt}{t\log(y/t)} + O(1) = \rho\log\left(\frac{\log JL}{\log L}\right) + O(1),$$

and thus the residual set has size $O(\frac{L\log L}{\log JL}\frac{x}{\log x})$. Taking $J = (\log x)^{1/3}$ and $L = (\log\log x)^{\rho-\varepsilon}$, the residual set has size at most $o(x/\log x) \leqslant (\rho/2-\varepsilon)\frac{x}{\log x}$, which gives (2.4), and so we're done.

2.2. **The Greedy Stage: Further details.** To successfully show (2.4) with y as large as $x(\log x)^\delta$, we use a hypergraph covering lemma of Pippenger-Spencer type introduced in [5]. This allows us to select residues b_q such that the sets

$$(S_{H^M} + b) \cap (b_q \bmod q) \cap [1, y]$$

are nearly disjoint.

It is convenient to separately consider the primes $q \in (z, x/2]$ in finer-than-dyadic blocks. Fix a real number $\xi > 1$ (which we will eventually take very close to 1) and define

$$\mathfrak{H} := \left\{H \in \{1, \xi, \xi^2, \dots\} : \frac{2y}{x} \leqslant H \leqslant \frac{y}{\xi z}\right\} \tag{2.6}$$

be the set of relevant scales H; we will consider those primes q in $(y/(\xi H), y/H]$ separately for each $H \in \mathfrak{H}$, noting that $\cup_{H\in\mathfrak{H}}(\frac{y}{\xi H}, \frac{y}{H}]$, is a subinterval of $(z, x/2]$. By (2.2) and (2.1) for $H \in \mathfrak{H}$ we have

$$2(\log x)^\delta \leqslant H \leqslant \frac{(\log x)^{1/2}}{\log\log x}. \tag{2.7}$$

For each $h \in \mathfrak{H}$, let $\mathcal{Q}_H$ be the set of primes $q \in (y/(\xi H), y/H]$ with $|I_q| \geqslant 1$. From (1.3), we have

$$|\mathcal{Q}_H| \sim \rho(1 - 1/\xi)\frac{y}{H\log x}. \tag{2.8}$$

Let

$$\mathcal{Q} = \bigcup_{H \in \mathfrak{H}} \mathcal{Q}_H.$$

For $q \in \mathcal{Q}$, let H_q be the unique element of $\mathfrak{H}$ such that

$$\frac{y}{\xi H_q} < q \leqslant \frac{y}{H_q}. \tag{2.9}$$

Now fix a real number M satisfying

$$4 + \delta < M \leqslant 5. \tag{2.10}$$

With H fixed, we will examine separately the effect of the sieving by primes in $[2, H^M]$ and by the primes in $(H^M, z]$. We denote by $\mathbf{b}$ a random residue class from $\mathbb{Z}/P\mathbb{Z}$, chosen with uniform probability, where we adopt the abbreviations

$$P = P(z), \quad \sigma = \sigma(z), \quad \mathbf{S} = S_z + \mathbf{b}$$

as well as the projections

$$P_1 = P(H^M), \quad \sigma_1 = \sigma(H^M), \quad \mathbf{b}_1 \equiv \mathbf{b} \pmod{P_1}, \quad \mathbf{S}_1 = S_{H^M} + \mathbf{b}_1 \tag{2.11}$$

and

$$P_2 = P(H^M, z), \quad \sigma_2 = \sigma(H^M, z), \quad \mathbf{b}_2 \equiv \mathbf{b} \pmod{P_2}, \quad \mathbf{S}_2 = S_{H^M,z} + \mathbf{b}_2 \tag{2.12}$$

with the convention that $\mathbf{b}_1 \in \mathbb{Z}/P_1\mathbb{Z}$ and $\mathbf{b}_2 \in \mathbb{Z}/P_2\mathbb{Z}$. Thus, $\mathbf{b}_1$ and $\mathbf{b}_2$ are each uniformly distributed, are independent of each other, and likewise $\mathbf{S}_1$ and $\mathbf{S}_2$ are independent. We also have the obvious relations

$$P = P_1P_2, \quad \sigma = \sigma_1\sigma_2, \quad \mathbf{S} = \mathbf{S}_1 \cap \mathbf{S}_2.$$

For prime q and $n \in \mathbb{Z}$, define the random set

$$\mathbf{AP}(J; q, n) := \{n + qh : 1 \leqslant h \leqslant J\} \cap \mathbf{S}_1 \tag{2.13}$$

that describes a portion of the progression $n \pmod q$ that survives the sieving process up to H^M. Let $K \geqslant 2$ be a fixed integer parameter, which we will eventually take to be very large. Given $\mathbf{S}_1$, the probability that $\mathbf{AP}(KH; q, n) \subset \mathbf{S}_2$ is about $\sigma_2^{|\mathbf{AP}(KH;q,n)|}$, and if this occurs then removing the residue class $n \bmod q$ will remove an essentially maximal number of elements. Central to our argument is the weight function

$$\boldsymbol{\lambda}(H; q, n) := \begin{cases} \dfrac{1}{\sigma_2^{|\mathbf{AP}(KH;q,n)|}} & \text{if } \mathbf{AP}(KH; q, n) \subset \mathbf{S}_2, \\ 0 & \text{otherwise.} \end{cases} \tag{2.14}$$

Informally, $\boldsymbol{\lambda}(H; q, n)$ then isolates those n with the (somewhat unlikely) property that the portion $\mathbf{AP}(KH; q, n)$ of the arithmetic progression $n \bmod q$ that survives the sieving process up to H^M, in fact also survives the sieving process all the way up to z. The weight nearly exactly counteracts the probability of this event, so that we anticipate $\boldsymbol{\lambda}(H; q, n)$ to be about 1 on average over n. In addition, $\boldsymbol{\lambda}(H; q, n)$ is skewed to be large for those n with $\mathbf{AP}(KH; q, n)$ large. We will focus attention on those n satisfying

$$-Ky < n \leqslant y,$$

for outside this interval, if $q \in \mathcal{Q}_H$ then $\mathbf{AP}(KH; q, n)$ does not intersect the interval $[1, y]$ of primary interest.

Our aim is thus first select a random $\mathbf{b} \in \mathbb{Z}/P\mathbb{Z}$, and show that with high probability the random sets $\mathbf{S}_1$ and $\mathbf{S}_2$ behave as we expect for all scales $H \in \mathfrak{H}$. This implies that there is a good fixed choice $b \in \mathbb{Z}/P\mathbb{Z}$ where the (now deterministic) function $\lambda(H_q; q, n)$ is suitably concentrated on residue classes $n \bmod q$ which contain many elements in $S = S_z + b$, for all q in a suitable subset $\mathcal{Q}' \subseteq \mathcal{Q}$. In particular, this means that if we then select a residue class $n_q \bmod q$ randomly with probability proportional to $\lambda(H_q; q, n)$, this residue class will typically contain many elements of S, for any $q \in \mathcal{Q}'$.

This is now precisely the situation of our hypergraph covering lemma, which we can then apply essentially as a black box. (The lemma is a minor variation of the one used in [5] based on the "Rödl nibble" or "semi-random" method; the proof is given in the appendix.) The conclusion from the lemma allows us to deduce that there is a choice of residue classes $n_q \bmod q$ for $q \in \mathcal{Q}'$ which cover almost all of S. If we then choose $b \bmod P(z, x/2)$ such that $b = n_q \bmod q$ for all $q \in \mathcal{Q}'$ we then obtain (2.4), and hence the result.

The paper is organized as follows. Theorem 1 has previously been reduced to that of establishing (2.4). We will then reduce this task further to that of establishing Theorem 2 (Second reduction) in the next section. In turn, Theorem 2 will be reduced to Theorem 3 (Third reduction) in the following section. The final section is then dedicated to establishing Theorem 3.

3. Greedy sieving via Hypergraph covering

In this section we use our hypergraph covering lemma (Lemma 3.1, given below) to reduce the proof of Theorem 1 to the claim that there is a good choice of b for the initial sieving, which is given by Theorem 2 below.

Recall the definition (2.9) of H_q and that S is the set $S_z + b$ depending on b.

Theorem 2 (Second reduction)**.** *Fix M satisfying* (2.10)*, fix δ satisfying* (2.3)*, and suppose $\varepsilon > 0$ is fixed and sufficiently small. If x is large (with respect to M, ε) then there exists an integer b and a set $\mathcal{Q}' \subset \mathcal{Q}$ such that*

(i) *one has*

$$|S \cap [1, y]| \leqslant 2\sigma y, \tag{3.1}$$

(ii) *for all $q \in \mathcal{Q}'$, one has*

$$\sum_{-Ky < n \leqslant y} \lambda(H_q; q, n) = \left(1 + O\left(\frac{1}{(\log x)^{\delta(1+\varepsilon)}}\right)\right)(K+1)y, \tag{3.2}$$

(iii) *for all but at most $\frac{\rho x}{8 \log x}$ elements n of $S \cap [1, y]$, one has*

$$\sum_{q \in \mathcal{Q}'} \sum_{h \leqslant KH_q} \lambda(H_q; q, n - qh) = \left(C_2 + O\left(\frac{1}{(\log x)^{\delta(1+\varepsilon)}}\right)\right)(K+1)y \tag{3.3}$$

for some quantity C_2 independent of n with

$$10^{2\delta} \leqslant C_2 \leqslant 100. \tag{3.4}$$

Theorem 2 is saying that there is a good choice of $b \in \mathbb{Z}/P\mathbb{Z}$ such that we can then perform the second sieving stage effectively. The conclusions are what we would expect for "typical" b, so this merely sets the stage for the greedy sieve.

If we remove a residue class $\mathbf{n_q} \bmod q$ where $\mathbf{n_q}$ is chosen randomly proportional to $\lambda(H_q; q, \cdot)$, then together (3.2) and (3.3) say that the expected number of times $n \in S \cap [1, y]$ is removed is about $C_2 > 1$ (apart from a small exceptional set of n). This means that if we could realize these random variables so that the behavior was very close to this expectation, we would sieve in a perfectly uniform manner and would successfully remove almost all of $S \cap [1, y]$. The fact that we can pass from the random variables to such a uniform sieve is a consequence of the hypergraph covering lemma. It is vital that $C_2 > 1$, and the fact that we will ultimately succeed with C_2 bounded (rather than of size $\log\log x$) corresponds to us being able to take y as large as $x(\log x)^{\delta}$.

The fact that we have good error terms in the asymptotics and the slightly stronger lower bound $C_2 > 10^{2\delta}$ is needed for our hypergraph covering lemma, but this is not a limiting feature of our argument.

Another way to look at Theorem 2 is that equation (3.2) says that $\lambda(H_q; q, n)$ is about 1 on average. However, when n is drawn from the smaller set $S \cap [1, y]$ (which has density $\approx \sigma$ in $[1, y]$), the quantity $\lambda(H_q; q, n - qh)$ appearing in (3.3) is biased to be a bit larger (in our construction, it will eventually behave like $\frac{\log y}{\log(y/q)}$ on the average over $q \in \mathcal{Q}'$), since $n \in AP(KH; q, n - hq)$ is already known to lie in S. It is this bias that ultimately allows us to gain somewhat over the trivial bound of $\gg x$ on the gap size in Theorem 1.

To reduce Theorem 1 to Theorem 2, we will use the following hypergraph covering lemma.

Lemma 3.1 (Hypergraph covering lemma)**.** *Suppose that $0 < \delta \leqslant \frac{1}{2}$, let $y \geqslant y_0(\delta)$ with $y_0(\delta)$ sufficiently large, and let V be finite set with $|V| \leqslant y$. Let $1 \leqslant s \leqslant y$, and suppose that $\mathbf{e}_1, \dots, \mathbf{e}_s$ are random subsets of V satisfying the following:*

$$|\mathbf{e}_i| \leqslant \frac{(\log y)^{1/2}}{\log\log y} \qquad (1 \leqslant i \leqslant s), \tag{3.5}$$

$$\mathbb{P}(v \in \mathbf{e}_i) \leqslant y^{-1/2-1/100} \qquad (v \in V, 1 \leqslant i \leqslant s), \tag{3.6}$$

$$\sum_{i=1}^{s} \mathbb{P}(v, v' \in \mathbf{e}_i) \leqslant y^{-1/2} \qquad (v, v' \in V, v \neq v'), \tag{3.7}$$

$$\left| \sum_{i=1}^{s} \mathbb{P}(v \in \mathbf{e}_i) - C_2 \right| \leqslant \eta \qquad (v \in V), \tag{3.8}$$

where C_2 and η satisfy

$$10^{2\delta} \leqslant C_2 \leqslant 100, \qquad \eta \geqslant \frac{1}{(\log y)^{\delta} \log\log y}. \tag{3.9}$$

Then there are subsets e_i of V, $1 \leqslant i \leqslant s$, with e_i being in the support of $\mathbf{e}_i$ for every i, and such that

$$\Big| V \setminus \bigcup_{i=1}^{s} e_i \Big| \leqslant C_3 \eta |V|, \tag{3.10}$$

where C_3 is an absolute constant.

This lemma is proven using almost exactly the same argument used to prove [5, Corollary 4] (after some minor changes of notation); we defer the proof to the appendix.

The conditions (3.5), (3.6) and (3.7) should be thought of as conditions which ensure that the randoms sets $\mathbf{e}_i$ typically spread out and cover most vertices in V fairly evenly. The condition (3.9) ensures that typically all vertices are covered slightly more than once in a uniform manner. Provided these conditions are fulfilled then the conclusion (3.10) is that there is a non-zero probability that virtually all vertices are covered, and so there is a deterministic realization of the random variables which covers virtually all the vertices. The key point is that C_2 can be taken to be bounded, since this means that the covering sets e_i are close to disjoint, and this is what allows us to improve the situation of trying to sieve independently for each q.

Reduction of Theorem 1 to Theorem 2. We are now in a position to deduce (2.4), and hence Theorem 1, from Theorem 2. Let b and $\mathcal{Q}'$ be the quantities whose existence is asserted by Theorem 2, and so $S = S_z + b$.

Property (iii) of Theorem 2 implies that there is a set $V \subseteq S \cap [1, y]$, containing all but at most $\frac{\rho x}{8 \log x}$ elements of $S \cap [1, y]$, and such that (3.3) holds for all $n \in V$. For each $q \in \mathcal{Q}'$, we choose a random integer $\mathbf{n}_q$ with probability density function

$$\mathbb{P}(\mathbf{n}_q = n) = \frac{\lambda(H_q; q, n)}{\sum_{-Ky < n' \leqslant y} \lambda(H_q; q, n')}. \tag{3.11}$$

Note that by (3.2) that the denominator is non-zero, so that this is a well-defined probability distribution. We will not need to assume any independence hypotheses on the $\mathbf{n}_q$. For each $q \in \mathcal{Q}'$, we then define the random subset $\mathbf{e}_q$ of V by the formula

$$\mathbf{e}_q := V \cap \{\mathbf{n}_q + hq : 1 \leqslant h \leqslant KH_q\}. \tag{3.12}$$

Our goal is to show that there are choices n_q of the random variable $\mathbf{n}_q$ which occur with positive probability such that the corresponding sets e_q cover most of V. Specifically, we wish to use Lemma 3.1 to show that

$$\Big| V \setminus \bigcup_{q \in \mathcal{Q}'} e_q \Big| \leqslant \frac{\rho x}{8 \log x}. \tag{3.13}$$

By construction, if (3.13) holds then for each $q \in \mathcal{Q}'$ there is a number n_q such that

$$e_q \subset \{n \in V : n \equiv n_q \pmod q\}.$$

Taking $b \equiv n_q \pmod q$ for all $q \in \mathcal{Q}'$, we find that

$$\big|(S_{x/2} + b) \cap [1, y]\big| \leqslant |S \cap [1, y] \setminus V| + \Big| V \setminus \bigcup_{q \in \mathcal{Q}'} e_q \Big| \leqslant \frac{\rho x}{8 \log x} + \frac{\rho x}{8 \log x} = \frac{\rho x}{4 \log x},$$

as required for (2.4). The fractions $\frac{1}{8}$ and $\frac{1}{4}$ above are irrelevant to the determination of the best exponent in Theorem 1, and were chosen for convenience.

Thus it remains to construct e_q satisfying (3.13), and this is accomplished by Lemma 3.1. We wish to apply Lemma 3.1 with $s = |\mathcal{Q}'|$, $\{\mathbf{e}_1, \dots, \mathbf{e}_s\} = \{\mathbf{e}_q : q \in \mathcal{Q}'\}$, C_2 as given by Theorem 2, and

$$\eta = \frac{\rho/20}{C_3 (\log x)^\delta}.$$

With this choice of parameters we see from (3.1), (1.2), and (2.1) that

$$C_3\eta|V| \leqslant \frac{\rho/10}{(\log x)^\delta}\frac{y}{\log z} \sim (\rho/10)\frac{x}{\log x}.$$

Hence, (3.13) follows from (3.10) if x is large enough. Thus, it suffices to verify the hypotheses (3.5), (3.6), (3.7), (3.8) and (3.9) of the lemma, which we accomplish using the conclusions (3.2) and (3.3) of Theorem 2.

Note that if $q \in \mathcal{Q}'$, then from (3.12) and (2.6) we have

$$|\mathbf{e}_q| \leqslant H_q \leqslant \frac{y}{z} = \frac{(\log x)^{1/2}}{\log\log x} \leqslant \frac{(\log y)^{1/2}}{\log\log y}$$

which gives (3.5). Similarly, for $n \in V$ and $q \in \mathcal{Q}'$, we have from (3.12), (3.11), and (2.14) that

$$\begin{aligned}\mathbb{P}(n \in \mathbf{e}_q) &= \sum_{1\leqslant h\leqslant KH_q} \mathbb{P}(\mathbf{n}_q = n - hq)\\ &\ll \frac{1}{y}\sum_{1\leqslant h\leqslant KH_q} \lambda(H_q; q, n-hq)\\ &\ll \frac{1}{y}H_q\sigma_2^{-H_q} \ll \frac{1}{y^{9/10}}\end{aligned}$$

which gives (3.6) for y large enough.

Applying (3.12), (3.11), (3.2), and (3.3) successively yields

$$\begin{aligned}\sum_{q\in\mathcal{Q}'}\mathbb{P}(v \in \mathbf{e}_q) &= \sum_{q\in\mathcal{Q}'}\sum_{h\leqslant KH_q}\mathbb{P}(\mathbf{n}_q = v - hq)\\ &= \sum_{q\in\mathcal{Q}'}\sum_{h\leqslant KH_q}\frac{\lambda(H_q; q, v-hq)}{\sum_n \lambda(H_q; q, n)}\\ &= C_2 + O((\log x)^{-\delta-\varepsilon}),\end{aligned}$$

and (3.8) follows. We now turn to (3.7). Observe from (3.12) that for distinct $v, v' \in V$, one can only have $v, v' \in \mathbf{e}_q$ if q divides $v - v'$. Since $|v - v'| \leqslant 2y$ and $q \geqslant z > \sqrt{2y}$, there is at most one q for which this is the case, and (3.7) now follows from (3.6). This concludes the derivation of (2.4) from Theorem 2. □

To complete the proof of Theorem 1, we need to prove Theorem 2 and Lemma 3.1. The proof of Theorem 2 depends on various first and second moment estimations of the weights, which are given in the next two sections. The proof of Lemma 3.1 will occupy the Appendix.

4. Concentration of $\lambda(H; q, n)$

In this section, we deduce Theorem 2 from the following moment calculations.

Theorem 3 (Third reduction)**.** *Assume that $M \geqslant 2$. Then*

(i) *One has*

$$\mathbb{E}|\mathbf{S}\cap[1,y]| = \sigma y, \tag{4.1}$$

$$\mathbb{E}|\mathbf{S}\cap[1,y]|^2 = \left(1+O\left(\frac{1}{\log y}\right)\right)(\sigma y)^2. \tag{4.2}$$

(ii) *For every $H\in\mathfrak{H}$, and for $j\in\{0,1,2\}$ we have*

$$\mathbb{E}\sum_{q\in\mathcal{Q}_H}\left(\sum_{-Ky<n\leqslant y}\boldsymbol{\lambda}(H;q,n)\right)^j = \left(1+O\left(\frac{1}{H^{M-2}}\right)\right)((K+1)y)^j|\mathcal{Q}_H|. \tag{4.3}$$

(iii) *For every $H\in\mathfrak{H}$, and for $j\in\{0,1,2\}$ we have*

$$\mathbb{E}\sum_{n\in\mathbf{S}\cap[1,y]}\left(\sum_{q\in\mathcal{Q}_H}\sum_{h\leqslant KH}\boldsymbol{\lambda}(H;q,n-qh)\right)^j = \left(1+O\left(\frac{1}{H^{M-2}}\right)\right)\left(\frac{|\mathcal{Q}_H|KH}{\sigma_2}\right)^j\sigma y. \tag{4.4}$$

We remind the reader that in Theorem 3 the random variables $\mathbf{S}$ and $\boldsymbol{\lambda}$ are defined in terms of the random variable $\mathbf{b}$ chosen uniformly in $\mathbb{Z}/P\mathbb{Z}$, not the random variables $\mathbf{n_q}$ we encountered in the previous section.

Note that for every $n\in[1,y]$ and $h\leqslant KH$ we have $n-qh\in[-Ky,y]$, so the quantity in (4.4) is well-defined. As with the previous theorem, the quantity $\boldsymbol{\lambda}(H;q,n)$ behaves like 1 on the average when n is drawn from $[-Ky,y]\cap\mathbb{Z}$, but for n drawn from $\mathbf{S}\cap[1,y]$ (in particular, $n\in\mathbf{S}_2$), the quantity $\boldsymbol{\lambda}(H;q,n-qh)$ is now biased to have an average value of approximately σ_2^{-1} because $n-qh+qh=n$ is automatically in $\mathbf{S}_2$; recall the definition (2.14) of $\lambda(H;q,n-qh)$.

Deduction of Theorem 2 from Theorem 3. We draw $\mathbf{b}$ uniformly at random from $\mathbb{Z}/P\mathbb{Z}$. It will suffice to generate a random set $\boldsymbol{\mathcal{Q}}'$ such that the random function $\boldsymbol{\lambda}$ defined in (2.14) satisfies the conclusions of Theorem 2 (with b replaced by $\mathbf{b}$) hold with positive probability - in fact, we will show that they hold with probability $1+o(1)$.

Assume that M satisfies (2.10). From Theorem 3(i) we have

$$\mathbb{E}\big||\mathbf{S}\cap[1,y]|-\sigma y\big|^2 \ll \frac{(\sigma y)^2}{\log y}.$$

Hence by Chebyshev's inequality, we see that

$$\mathbb{P}\left(|\mathbf{S}\cap[1,y]|\leqslant 2\sigma y\right) = 1-O(1/\log x), \tag{4.5}$$

verifying (3.2) in Theorem 2. Let $H\in\mathfrak{H}$. From Theorem 3(ii) we have (recall that our implied constants may depend on K)

$$\mathbb{E}\sum_{q\in\mathcal{Q}_H}\left(\sum_{-Ky<n\leqslant y}\boldsymbol{\lambda}(H;q,n)-(K+1)y\right)^2 \ll \frac{y^2|\mathcal{Q}_H|}{H^{M-2}}. \tag{4.6}$$

Now let $\boldsymbol{\mathcal{Q}}'_H$ be the subset of $q\in\mathcal{Q}_H$ with the property that

$$\Big|\sum_{-Ky<n\leqslant y}\boldsymbol{\lambda}(H;q,n)-(K+1)y\Big| \leqslant \frac{y}{H^{1+\varepsilon}}. \tag{4.7}$$

It follows from (4.6) and (4.7) that

$$\mathbb{E}|\mathcal{Q}_H\backslash\boldsymbol{\mathcal{Q}}'_H| \ll \frac{|\mathcal{Q}_H|}{H^{M-4-2\varepsilon}}. \tag{4.8}$$

By Markov's inequality, it follows that with probability $1-O(H^{-\varepsilon})$, one has

$$|\mathcal{Q}_H\backslash\boldsymbol{\mathcal{Q}}'_H| \ll \frac{|\mathcal{Q}_H|}{H^{M-4-3\varepsilon}}. \tag{4.9}$$

By (2.10), we have $M > 4+3\varepsilon$ for small enough ε, that is, the exponent in the denominator in (4.9) is positive. Since $\sum_{H\in\mathfrak{H}} H^{-\varepsilon} \ll (y/x)^{-\varepsilon} \ll (\log x)^{-\delta\varepsilon}$, with probability $1-O((\log x)^{-\delta\varepsilon})$ the relation (4.9) holds for every $H\in\mathfrak{H}$ simultaneously. We now set

$$\boldsymbol{\mathcal{Q}}' := \bigcup_{H\in\mathfrak{H}} \boldsymbol{\mathcal{Q}}'_H.$$

Then, on the probability $1-o(1)$ event that (4.9) holds for every H and that (4.5) holds, items (i) (3.1) and (ii) (3.2) of Theorem 2 follow upon recalling (4.7) and the lower bound $H \gg (\log x)^{\delta}$.

We work on part (iii) of Theorem 2 using Theorem 3(iii) in a similar fashion to previous arguments. We have

$$\mathbb{E}\sum_{n\in\mathbf{S}\cap[1,y]}\left|\sum_{q\in\mathcal{Q}_H}\sum_{h\leqslant KH}\boldsymbol{\lambda}(H;q,n-qh)-\frac{|\mathcal{Q}_H|KH}{\sigma_2}\right|^2 \ll \frac{1}{H^{M-2}}\left(\frac{|\mathcal{Q}_H|KH}{\sigma_2}\right)^2\sigma y.$$

If we let $\mathcal{E}_H$ denote the set of $n\in\mathbf{S}\cap[1,y]$ such that

$$\left|\sum_{q\in\mathcal{Q}_H}\sum_{h\leqslant KH}\boldsymbol{\lambda}(H;q,n-qh)-\frac{|\mathcal{Q}_H|KH}{\sigma_2}\right| \geqslant \frac{|\mathcal{Q}_H|KH}{\sigma_2 H^{(M-2)/2-\varepsilon}}, \tag{4.10}$$

then

$$\mathbb{E}|\mathcal{E}_H| \ll \frac{\sigma y}{H^{\varepsilon}}.$$

By Markov's inequality, we conclude that $|\mathcal{E}_H| \leqslant \sigma y/H^{\varepsilon/2}$ with probability $1-O(H^{-\varepsilon/2})$.

We next estimate the contribution from "bad" primes $q\in\mathcal{Q}_H\backslash\boldsymbol{\mathcal{Q}}'_H$. For any $h\leqslant H$, by Cauchy-Schwarz we have

$$\mathbb{E}\sum_{n\in\mathbf{S}\cap[1,y]}\sum_{q\in\mathcal{Q}_H\backslash\boldsymbol{\mathcal{Q}}'_H}\boldsymbol{\lambda}(H;q,n-hq) \leqslant \big(\mathbb{E}|\mathcal{Q}_H\backslash\boldsymbol{\mathcal{Q}}'_H|\big)^{1/2}\left(\mathbb{E}\sum_{q\in\mathcal{Q}_H\backslash\boldsymbol{\mathcal{Q}}'_H}\Big|\sum_{n=1}^{y}\boldsymbol{\lambda}(H;q,n-hq)\Big|^2\right)^{1/2}$$

and by the triangle inequality, (4.6) and (4.8),

$$\begin{aligned}\mathbb{E}\sum_{q\in\mathcal{Q}_H\backslash\boldsymbol{\mathcal{Q}}'_H}\Big|\sum_{n=1}^{y}\boldsymbol{\lambda}(H;q,n-hq)\Big|^2 &\leqslant 2\mathbb{E}\sum_{q\in\mathcal{Q}_H\backslash\boldsymbol{\mathcal{Q}}'_H}\left(\Big|\sum_{n=1}^{y}\boldsymbol{\lambda}(H;q,n-hq)-(K+1)y\Big|^2+(K+1)^2y^2\right)\\ &\ll \frac{y^2|\mathcal{Q}_H|}{H^{M-4-2\varepsilon}}.\end{aligned}$$

Therefore, by (4.8) and summing over $h \leqslant KH$,

$$\mathbb{E} \sum_{n\in\mathbf{S}\cap[1,y]} \sum_{q\in\mathcal{Q}_H\setminus\boldsymbol{\mathcal{Q}}'_H} \sum_{h\leqslant KH} \boldsymbol{\lambda}(H;q,n-hq) \ll \frac{y|\mathcal{Q}_H|}{H^{M-5-2\varepsilon}}.$$

Let $\mathcal{E}'_H$ denote the set of $n \in \mathbf{S}\cap[1,y]$ so that

$$\sum_{q\in\mathcal{Q}_H\setminus\boldsymbol{\mathcal{Q}}'_H} \sum_{h\leqslant KH} \boldsymbol{\lambda}(H;q,n-hq) \geqslant \frac{|\mathcal{Q}_H|KH}{H^{(1+\varepsilon)\delta}\sigma_2}. \tag{4.11}$$

Then

$$\mathbb{E}|\mathcal{E}'_H| \ll \frac{yH^{\delta(1+\varepsilon)}\sigma_2}{H^{M-4-2\varepsilon}} \ll \sigma y \frac{\log H}{H^{M-4-\delta-3\varepsilon}}.$$

By Markov's inequality, $|\mathcal{E}'_H| \leqslant \sigma y/H^{\varepsilon}$ with probability $1-O(1/H^{M-4-\delta-5\varepsilon})$. By (2.10) again, if ε is small enough then $M-4-\delta-5\varepsilon > \varepsilon$. Consider the event that (4.5) holds, and that for every H, we have (4.9), $|\mathcal{E}_H| \leqslant \sigma y/H^{\varepsilon/2}$ and $|\mathcal{E}'_H| \leqslant \sigma y/H^{1+\varepsilon}$. This simultaneous event happens with positive probability on account of $\sum_{H\in\mathfrak{H}} H^{-\eta} \ll (\log x)^{-\delta\eta}$ for any $\eta>0$. As mentioned before, items (i) and (ii) of Theorem 2 hold. Now let

$$\mathcal{N} = \mathbf{S}\cap[1,y]\setminus \bigcup_{H\in\mathfrak{H}} (\mathcal{E}_H \cup \mathcal{E}'_H).$$

The number of exceptional elements satisfies

$$\left| \bigcup_{H\in\mathfrak{H}} (\mathcal{E}_H \cup \mathcal{E}'_H) \right| \ll \frac{\sigma y}{(\log x)^{\delta(1+\varepsilon)}},$$

which is smaller than $\frac{\rho x}{8\log x}$ for large x. It remains to verify (3.3) for $n\in\mathcal{N}$. Since $n\notin\mathcal{E}_H$ and $n\notin\mathcal{E}'_H$ for every H, the inequalities opposite to those in (4.10) and (4.11) hold, and we have for each $H\in\mathfrak{H}$ the asymptotic

$$\sum_{q\in\mathcal{Q}'_H} \sum_{h\leqslant KH} \boldsymbol{\lambda}(H;q,n-qh) = \left(1+O\left(\frac{1}{H^{(1+\varepsilon)\delta}}\right)\right)\frac{|\mathcal{Q}_H|KH}{\sigma_2}.$$

Therefore,

$$\begin{aligned}\sum_{q\in\boldsymbol{\mathcal{Q}}'} \sum_{h\leqslant KH_q} \boldsymbol{\lambda}(H_q;q,n-qh) &= \sum_{H\in\mathfrak{H}} \sum_{q\in\boldsymbol{\mathcal{Q}}'_H} \sum_{h\leqslant KH} \boldsymbol{\lambda}(H;q,n-qh)\\ &= \left(1+O\left(\frac{1}{(\log x)^{(1+\varepsilon)\delta}}\right)\right) C_2\times(K+1)y\end{aligned}$$

where

$$C_2 := \frac{K}{(K+1)y} \sum_{H\in\mathfrak{H}} \frac{|\mathcal{Q}_H|H}{\sigma_2}.$$

This verifies (3.3). From (2.8), we see that C_2 does not depend on n (C_2 depends only on x). Using (1.2) and (2.8),

$$C_2 \sim \frac{K}{(K+1)y}\rho(1-1/\xi) \sum_{H\in\mathfrak{H}} \frac{\log z}{\log(H^M)}\ \frac{yH}{H\log x} \qquad (x\to\infty).$$

Recalling the definitions (2.1) of y and (2.2) of z, together with the bounds (2.7) on H, we thus have as $x \to \infty$,

$$\begin{aligned} C_2 &\sim \frac{K\rho(1-1/\xi)}{M(K+1)} \sum_{H \in \mathfrak{H}} \frac{1}{\log H} \\ &= \frac{K\rho(1-1/\xi)}{M(K+1)} \sum_{2(\log x)^\delta \leqslant \xi^j \leqslant \xi^{-1}(\log x)^{1/2}/\log\log x} \frac{1}{j \log \xi}. \end{aligned}$$

Summing on j we conclude that

$$C_2 \sim \frac{K\rho}{M(K+1)} \frac{1-1/\xi}{\log \xi} \log\left(\frac{1}{2\delta}\right).$$

Finally, recalling (2.3), we see that if K is large enough, ξ is sufficiently close to 1 and M sufficiently close to $4+\delta$, then

$$C_2 \geqslant 10^{2\delta},$$

as required for (3.4). □

Remark 9. The limit our methods appears to be an exponent $e^{-1/\rho} - o(1)$ in Theorem 1. Such a bound assumes that we may succeed with the previous argument for any choice of $M > 1$, any $C_2 > 1$ and with $z = y/(\log x)^{1+o(1)}$ in place of of $z = y/(\log x)^{1/2+o(1)}$. Then the above calculation reveals that $C_2 > 1$ provided $\rho \log(1/\delta) > 1$. Each of these conditions appears to be essential in the succeeding arguments in the next sections.

It remains to establish Theorem 3. This is the objective of the next section of the paper.

5. Computing correlations

In this section, we verify the claims in Theorem 3. We will frequently need to compute k-point correlations of the form

$$\mathbb{P}\left(n_1, \dots, n_k \in \mathbf{S}_2\right)$$

for various integers $n_1, \dots, n_k$ (not necessarily distinct). Heuristically, since $\mathbf{S}_2$ avoids I_p residue classes modulo p for each p, we expect that the above probability is roughly σ_2^k for typical choices of $n_1, \dots, n_k$. Unfortunately, there is some fluctuation from this prediction, most obviously when two or more of the $n_1, \dots, n_k$ are equal, but also if the reductions $n_i \pmod p, n_j \pmod p$ for some prime $p \in (H^M, z]$ have the same difference as two elements of I_p. Fortunately we can control these fluctuations to be small on average. To formalize this statement we need some notation. Let $\mathcal{D}_H \subset \mathbb{N}$ denote the collection of squarefree numbers d, all of whose prime factors lie in $(H^M, z]$. This set includes 1, but we will frequently remove 1 and work instead with $\mathcal{D}_H \backslash \{1\}$. For each $d \in \mathcal{D}_H$, let $I_d \subset \mathbb{Z}/d\mathbb{Z}$ denote the collection of residue classes $a \bmod d$ such that $a \bmod p \in I_p$ for all $p \mid d$. Recall the defnition of the difference set $\mathcal{A} - \mathcal{B} := \{a - b : a \in \mathcal{A}, b \in \mathcal{B}\}$. For any integer m and any parameter $A > 0$, we define the error function

$$E_A(m; H) := \sum_{d \in \mathcal{D}_H \backslash \{1\}} \frac{A^{\omega(d)}}{d} 1_{m \,(\mathrm{mod}\ d) \in I_d - I_d}, \tag{5.1}$$

where $\omega(d)$ is the number of prime factors of d. The quantity $E_A(m;H)$ looks complicated, but in practice it will be quite small on average over m. We also observe that E_A is an even function: $E_A(-m;H) = E_A(m;H)$.

Before we start our proof of Theorem 3, we first need two preparatory lemmas. The following lemmas hold for general H, not necessarily restricted to $H \in \mathfrak{H}$. Recall that implied constants in $O-$ may depend on B and M.

Lemma 5.1. *Let $10 < H < z^{1/M}$, $1 \leqslant \ell \leqslant 10KH$, and suppose that $\mathcal{U} \subset \mathcal{V}$ are finite sets of integers with $|\mathcal{V}| = \ell$. Then we have*

$$\mathbb{P}(\mathcal{U} \subset \mathbf{S}_2) = \sigma_2^{|\mathcal{U}|}\left(1 + O\left(\frac{|\mathcal{U}|^2}{H^M}\right) + O\left(\frac{1}{\ell^2}\sum_{\substack{v,v'\in\mathcal{V}\\ v\neq v'}} E_{2\ell^2 B}(v-v';H)\right)\right).$$

Remark 10. The numbers in $\mathcal{V}\setminus\mathcal{U}$ are "dummy variables", but it is often convenient to include them. Typically, $\mathcal{U}$ will be an irregular subset, with unknown size, of a regular set $\mathcal{V}$, whose size is known. We often have better control of the error averaged over the larger set.

Proof. For each prime $p \in (H^M, z]$, let $\mathbf{b}_{2,p} \in \mathbb{Z}/p\mathbb{Z}$ be the reduction of $\mathbf{b}_2$ modulo p, thus each $\mathbf{b}_{2,p}$ is uniformly distributed in $\mathbb{Z}/p\mathbb{Z}$ and the $\mathbf{b}_{2,p}$ are independent in p. Let N_p denote the set of residue classes $\mathcal{U} \pmod p$. By the Chinese Remainder Theorem, we thus have

$$\begin{aligned}\mathbb{P}(\mathcal{U} \subset \mathbf{S}_2) &= \prod_{p\in(H^M,z]} \mathbb{P}(N_p \cap (\mathbf{b}_{2,p} + I_p) = \emptyset)\\ &= \prod_{p\in(H^M,z]} (1 - \mathbb{P}(\mathbf{b}_{2,p} \in N_p - I_p))\\ &= \prod_{p\in(H^M,z]} \left(1 - \frac{|N_p - I_p|}{p}\right).\end{aligned}$$

Let $k = |\mathcal{U}|$. We may crudely estimate the size of the difference set $N_p - I_p$ by

$$k|I_p| \geqslant |N_p - I_p| \geqslant k|I_p| - |I_p| \sum_{u,u'\in\mathcal{U}, u\neq u'} 1_{u-u' \,(\mathrm{mod}\ p)\in I_p - I_p}.$$

Since $|I_p| \leqslant B$ and $k \leqslant 10H$, we have $k|I_p| < 10KBH < p/10$ for x large enough in terms of M. Thus,

$$\left(1 - \frac{|N_p - I_p|}{p}\right) = \left(1 - \frac{k|I_p|}{p}\right)\left(1 + \frac{k|I_p| - |N_p - I_p|}{p - k|I_p|}\right) = \left(1 - \frac{k|I_p|}{p}\right)\Delta_p,$$

where

$$\begin{aligned}1 \leqslant \Delta_p &\leqslant 1+\frac{2B}{p}\sum_{u,u'\in\mathcal{U},u\neq u'} 1_{u-u' \,(\mathrm{mod}\ p)\in I_p-I_p}\\ &\leqslant \prod_{u,u'\in\mathcal{U},u\neq u'} \exp\left\{2B\frac{1_{u-u' \,(\mathrm{mod}\ p)\in I_p-I_p}}{p}\right\}\\ &\leqslant \prod_{v,v'\in\mathcal{V},v\neq v'} \exp\left\{2B\frac{1_{v-v' \,(\mathrm{mod}\ p)\in I_p-I_p}}{p}\right\}.\end{aligned}$$

Here we have enlarged the summation over pairs of numbers from $\mathcal{V}$. We have

$$\prod_{H^M<p\leqslant z}\left(1-\frac{k|I_p|}{p}\right)=\sigma_2^k\left(1+O\left(\frac{k^2}{H^M}\right)\right).$$

By the arithmetic mean-geometric mean inequality, we have

$$\begin{aligned}\prod_{p\in(H^M,z]}\Delta_p &\leqslant \prod_{v,v'\in\mathcal{V},v\neq v'}\ \prod_{p\in(H^M,z]}\exp\left\{2B\frac{1_{v-v' \,(\mathrm{mod}\ p)\in I_p-I_p}}{p}\right\}\\ &\leqslant \frac{2}{\ell^2-\ell}\sum_{v,v'\in\mathcal{V},v\neq v'}\ \prod_{p\in(H^M,z]}\exp\left\{2B\left(\frac{\ell^2-\ell}{2}\right)\frac{1_{v-v' \,(\mathrm{mod}\ p)\in I_p-I_p}}{p}\right\}\\ &\leqslant \frac{2}{\ell^2-\ell}\sum_{v,v'\in\mathcal{V},v\neq v'}\ \prod_{p\in(H^M,z]}\left(1+2B\ell^2\frac{1_{v-v' \,(\mathrm{mod}\ p)\in I_p-I_p}}{p}\right).\end{aligned}$$

Recalling the definition (5.1) of $E_A(n;H)$ we see that

$$\begin{aligned}\prod_{p\in(H^M,z]}\Delta_p &\leqslant \frac{2}{\ell^2-\ell}\sum_{v,v'\in\mathcal{V},v\neq v'}\big(1+E_{2B\ell^2}(v-v';H)\big)\\ &= 1+\frac{2}{\ell^2-\ell}\sum_{v,v'\in\mathcal{V},v\neq v'}E_{2B\ell^2}(v-v';H).\end{aligned}$$

□

To estimate the average contribution of the errors $E_{2B\ell^2}(v-v')$ appearing in the above lemma, we will use the following estimate.

Lemma 5.2. *Suppose that $10<H<z^{1/M}$, and that $(m_t)_{t\in T}$ is a sequence of integers indexed by a finite set T, obeying the bounds*

$$\sum_{t\in T}1_{m_t\equiv a \ (\mathrm{mod}\ d)}\ll\frac{X}{\phi(d)}+R \tag{5.2}$$

for some $X,R>0$ and all $d\in\mathcal{D}_H\backslash\{1\}$ and $a\in\mathbb{Z}/d\mathbb{Z}$. Then, for any $0<A$ satisfying $AB^2\leqslant H^M$ and any integer j, one has

$$\sum_{t\in T}E_A(m_t+j;H)\ll X\frac{A}{H^M}+R\exp\left(AB^2\log\log y\right).$$

In practice, R will be much smaller than X, and the first term on the right-hand side will dominate.

Proof. From the Chinese Remainder Theorem and (1.1), we see that for any $d \in \mathcal{D}_H$, we have

$$|I_d| = \prod_{p \,|\, d} |I_p| \leqslant B^{\omega(d)}.$$

In particular, the difference set $I_d - I_d \subset \mathbb{Z}/d\mathbb{Z}$ obeys the bound

$$|I_d - I_d| \leqslant B^{2\omega(d)}.$$

From (5.1), (5.2) we thus have

$$\begin{aligned}\sum_{t\in T} E_A(m_t + j; H) &= \sum_{d\in\mathcal{D}_H\setminus\{1\}} \frac{A^{\omega(d)}}{d} \sum_{a\in I_d - I_d} \#\{t \in T : m_t + j \equiv a \pmod d\} \\ &\ll \sum_{d\in\mathcal{D}_H\setminus\{1\}} \frac{(AB^2)^{\omega(d)}}{d} \left(\frac{X}{\phi(d)} + R\right).\end{aligned}$$

From Euler products and Mertens' theorem (for primes) we have

$$\sum_{d\in\mathcal{D}_H} \frac{(AB^2)^{\omega(d)}}{d} = \prod_{p\in(H^M,z]} (1 + AB^2/p) \leqslant \exp\{AB^2 \log\log y\}$$

and

$$\sum_{d\in\mathcal{D}_H} \frac{(AB^2)^{\omega(d)}}{d\phi(d)} = \prod_{p\in(H^M,z]} \left(1 + \frac{AB^2}{p^2 - p}\right) \leqslant \exp\{AB^2/H^M\} \leqslant 1 + O(A/H^M). \qquad \square$$

Finally, we are now in a position to complete the proof of Theorem 3.

Proof of Theorem 3 (i). By linearity of expectation, we have

$$\mathbb{E}|\mathbf{S} \cap [1, y]| = \sum_{1\leqslant n\leqslant y} \mathbb{P}(n \in \mathbf{S}).$$

Since the set S is periodic with period P and has density σ, the summands here are all equal to σ, and (4.1) follows. Now we consider (4.2). Here we decompose $\mathbf{S}$ as $\mathbf{S} = \mathbf{S}_1 \cap \mathbf{S}_2$ using (2.11) and (2.12) with

$$H = \frac{1}{4}(\log y)^{1/M}.$$

By the Prime Number Theorem,

$$P_1 = \exp\{(1 + o(1))H^M\} \leqslant y^{1/4+o(1)}. \tag{5.3}$$

By linearity of expectation,

$$\begin{aligned}\mathbb{E}|\mathbf{S} \cap [1, y]|^2 &= \sum_{n_1,n_2\leqslant y} \mathbb{P}\left(n_1, n_2 \in \mathbf{S}\right) \\ &= \sum_{n_1,n_2\leqslant y} \mathbb{P}\left(n_1, n_2 \in \mathbf{S}_1\right) \mathbb{P}\left(n_1, n_2 \in \mathbf{S}_2\right).\end{aligned}$$

Observe that the probability $\mathbb{P}(n_1, n_2 \in \mathbf{S}_1)$ depends only on the reductions $\ell_1 :\equiv n_1 \pmod{P_1}$, $\ell_2 :\equiv n_2 \pmod{P_1}$. Also, applying Lemma 5.1 (with $\mathcal{U} = \mathcal{V} = \{n_1, n_2\}$), we have

$$\mathbb{P}(n_1, n_2 \in \mathbf{S}_2) = (1 + O(E_{8B}(n_1 - n_2; H)))\,\sigma_2^2.$$

Therefore,

$$\begin{aligned}\mathbb{E}|\mathbf{S} \cap [1, y]|^2 &= \sum_{1 \leqslant \ell_1, \ell_2 \leqslant P_1} \mathbb{P}(\ell_1, \ell_2 \in \mathbf{S}_1) \sum_{\substack{1 \leqslant n_1, n_2 \leqslant y \\ n_1 \equiv \ell_1 \pmod{P_1} \\ n_2 \equiv \ell_2 \pmod{P_1}}} \mathbb{P}(n_1, n_2 \in \mathbf{S}_2) \\ &= \sigma_2^2 \sum_{1 \leqslant \ell_1, \ell_2 \leqslant P_1} \mathbb{P}(\ell_1, \ell_2 \in \mathbf{S}_1) \left(\frac{y}{P_1} + O(1)\right)^2 + \\ &\quad + O\left(\sigma_2^2 \sum_{1 \leqslant \ell_1, \ell_2 \leqslant P_1} \mathbb{P}(\ell_1, \ell_2 \in \mathbf{S}_1) \sum_{\substack{1 \leqslant n_1, n_2 \leqslant y \\ n_1 \equiv \ell_1 \pmod{P_1} \\ n_2 \equiv \ell_2 \pmod{P_1}}} E_{8B}(n_1 - n_2; H)\right).\end{aligned} \tag{5.4}$$

By the definition (2.11),

$$\sum_{1 \leqslant \ell_1, \ell_2 \leqslant P_1} \mathbb{P}(\ell_1, \ell_2 \in \mathbf{S}_1) = \mathbb{E}\,|\mathbf{S}_1 \cap [1, P_1]|^2 = (\sigma_1 P_1)^2, \tag{5.5}$$

since $|\mathbf{S}_1 \cap [1, P_1]| = \sigma_1 P$ always. Next, fix $\ell_1, \ell_2 \in \mathbb{Z}/P_1\mathbb{Z}$. Direct counting shows that for any n_1, natural number $d \in \mathcal{D}_{H^+}$ and residue class $a \bmod d$, we have

$$\#\{n_2 \leqslant y : n_2 \equiv \ell_2 \pmod{P_1}, n_1 - n_2 \equiv a \pmod{d}\} \ll \frac{y}{dP_1} + 1 \leqslant \frac{y}{\phi(d)P_1} + 1.$$

Applying Lemma 5.2 to the inner sum over n_2, we deduce that

$$\begin{aligned}\sum_{\substack{1 \leqslant n_1, n_2 \leqslant y \\ n_1 \equiv \ell_1 \pmod{P_1} \\ n_2 \equiv \ell_2 \pmod{P_1}}} E_{8B}(n_1 - n_2; H) &\ll \left(\frac{y}{P_1}\right)^2 \frac{1}{H^M} + \frac{y}{P_1} \exp(O(\log\log y)) \\ &\ll \frac{y^2}{P_1^2 H^M} \ll \frac{y^2}{\log y}\end{aligned} \tag{5.6}$$

using (5.3). Inserting the bounds (5.5) and (5.6) into (5.4) completes the proof of (4.2). □

Proof of Theorem 3 (ii). Let $H \in \mathfrak{H}$. The case $j = 0$ is trivial, so we turn attention to the $j = 1$ claim:

$$\mathbb{E} \sum_{q \in \mathcal{Q}_H} \sum_{-Ky \leqslant n \leqslant y} \boldsymbol{\lambda}(H; q, n) = \left(1 + O\left(\frac{1}{H^{M-2}}\right)\right)(K+1)y|\mathcal{Q}_H|. \tag{5.7}$$

The left-hand expands as

$$\mathbb{E} \sum_{q \in \mathcal{Q}_H} \sum_{-Ky \leqslant n \leqslant y} \frac{1_{\mathbf{AP}(KH; q, n) \subset \mathbf{S}_2}}{\sigma_2^{|\mathbf{AP}(KH; q, n)|}}.$$

Recalling the splitting (2.11) and (2.12), that $\mathbf{b}_1$ and $\mathbf{b}_2$ are independent, and consequently that $\mathbf{AP}(KH; q, n)$ and $\mathbf{S}_2$ are independent (since the sets $\mathbf{AP}(KH; q, n)$ defined in (2.13) are determined by $\mathbf{S}_1$). The above expression then equals

$$\sum_{q\in\mathcal{Q}_H}\sum_{-Ky\leqslant n\leqslant y}\sum_{b_1}\frac{\mathbb{P}(\mathbf{b}_1=b_1)}{\sigma_2^{|\mathbf{AP}(KH;q,n)|}}\mathbb{P}(\mathbf{AP}(KH;q,n)\subset\mathbf{S}_2).$$

Fix $\mathbf{b}_1$ and apply Lemma 5.1 with $\mathcal{U}=\mathbf{AP}(KH;q,n)$ and $\mathcal{V}=\{n+qh : 1\leqslant h\leqslant KH\}$. We find that the left side of (5.7) equals

$$\sum_{q\in\mathcal{Q}_H}\sum_{-Ky\leqslant n\leqslant y}\left(1+O\left(\frac{1}{H^{M-2}}\right)+O\left(\frac{1}{H^2}\sum_{\substack{1\leqslant h,h'\leqslant KH\\ h\neq h'}}E_{2BK^2H^2}(qh-qh';H)\right)\right).$$

Clearly it suffices to show that

$$\sum_{q\in\mathcal{Q}_H}E_{2BK^2H^2}(qh-qh';H)\ll\frac{|\mathcal{Q}_H|}{H^{M-2}}$$

for any distinct h, h' satisfying $1\leqslant h, h'\leqslant KH$. For future reference we will show the more general estimate

$$\sum_{q\in\mathcal{Q}_H}E_{8BK^2H^2}(q\ell+k;H)\ll\frac{|\mathcal{Q}_H|}{H^{M-2}} \tag{5.8}$$

uniformly for any integer k and $0<|\ell|\leqslant KH$. Note that $E_A(n;H)$ is increasing in A.

To prove (5.8), fix ℓ, k. If $d\in\mathcal{D}_H\backslash\{1\}$ and $a \bmod d$ is a residue class, all the prime divisors of d are larger than $H^M > KH \geqslant |\ell|$; meanwhile, q is larger than z and is hence coprime to d. Thus the relation $q\ell\equiv a \pmod d$ only holds for q in at most one residue class modulo d, and hence by the Brun–Titchmarsh inequality we have

$$\#\{q\in\mathcal{Q}_H : q\ell\equiv a \pmod d\}\ll\frac{y/H}{\phi(d)\log y}$$

when (say) $d\leqslant\sqrt{y}$ (recall that $H\leqslant(\log y)^{1/2}$ by (2.7)). For $d>\sqrt{y}$, we discard the requirement that q be prime, and obtain the crude bound

$$\#\{q\in\mathcal{Q}_H : q\ell\equiv a \pmod d\}\ll\frac{y/H}{d}+1\leqslant\frac{y/H}{\sqrt{y}}.$$

Thus for all d we have

$$\#\{q\in\mathcal{Q}_H : q\ell\equiv a \pmod d\}\ll\frac{y}{H\phi(d)\log y}+\frac{\sqrt{y}}{H}$$

and hence by Lemma 5.2,

$$\begin{aligned}\sum_{q\in\mathcal{Q}_H}E_{8BK^2H^2}(q\ell+k;H)&\ll\frac{y}{H\log y}\frac{H^2}{H^M}+\frac{\sqrt{y}}{H}\exp\left(O(H^2\log\log y)\right)\\&\ll|\mathcal{Q}_H|H^{2-M}+\frac{\sqrt{y}}{H}\exp\left(O(H^2\log\log y)\right).\end{aligned}$$

We note that the O-bound in the exponential depends on B and K. The claim (5.8) now follows from the upper bound in (2.7), namely that $H \leqslant (\log y)^{1/2}(\log\log y)^{-1}$, together with the bounds (2.8) on $|\mathcal{Q}_H|$. Incidentally, this is the only part of the proof that requires the full strength of the upper bound in (2.7), but it does however constrain the size of z.

Now we turn to the $j = 2$ case of Theorem 3(ii), which is

$$\mathbb{E}\sum_{q\in\mathcal{Q}_H}\Big(\sum_{-Ky\leqslant n\leqslant y}\boldsymbol{\lambda}(H;q,n)\Big)^2 = \Big(1+O\Big(\frac{1}{H^{M-2}}\Big)\Big)(K+1)^2y^2|\mathcal{Q}_H|.$$

The left-hand side may be expanded as

$$\mathbb{E}\sum_{q\in\mathcal{Q}_H}\sum_{-Ky\leqslant n_1,n_2\leqslant y}\frac{1_{\mathbf{AP}(KH;q,n_1)\cup\mathbf{AP}(KH;q,n_2)\subset\mathbf{S}_2}}{\sigma_2^{|\mathbf{AP}(KH;q,n_1)|+|\mathbf{AP}(KH;q,n_2)|}}.$$

Apply Lemma 5.1 with

$$\begin{aligned}\mathcal{U} &= \mathbf{AP}(KH;q,n_1)\cup\mathbf{AP}(KH;q,n_2),\\ \mathcal{V} &= \{n_1+qh : 1\leqslant h\leqslant KH\}\cup\{n_2+qh : 1\leqslant H\leqslant KH\},\end{aligned}$$

so that $|\mathcal{V}| = \ell \geqslant KH$. Noting that $\mathbf{S}_2$ is independent of both $\mathbf{AP}(KH;q,n_1)$ and $\mathbf{AP}(KH;q,n_2)$, we may write the previous expression as

$$\sum_{q\in\mathcal{Q}_H}\sum_{-Ky\leqslant n_1,n_2\leqslant y}\Bigg(1+O\Big(\frac{1}{H^{M-2}}\Big)+O\Bigg(\frac{1}{H^2}\sum_{h,h'\leqslant KH}\Big(1_{h\neq h'}E_{8BK^2H^2}(qh-qh';H)+\\ +1_{n_1\neq n_2}E_{8BK^2H^2}(n_1+qh-n_2-qh';H)\Big)\Bigg)\Bigg).$$

Using (5.8), we obtain an acceptable main term and error terms for everything except for the summands with $h = h'$. For any fixed n_2, any $d \geqslant 1$ and $a \bmod d$,

$$\#\{-Ky\leqslant n_1\leqslant y : n_1-n_2\equiv a \pmod d\}\ll\frac{y}{d}+1$$

so by Lemma 5.2, we have

$$\sum_{-Ky\leqslant n_1,n_2\leqslant y}E_{8BK^2H^2}(n_1-n_2;H)\ll y^2\frac{H^2}{H^M}+y\exp\big(O(H^2\log\log y)\big)\ll\frac{y^2}{H^{M-2}},$$

again using (2.7). This completes the proof of the $j = 2$ case, and so we have established (4.3). □

Proof of Theorem 3(iii). The $j = 0$ case follows from the $j = 1$ case of part (i) (that is, (4.2)), so we turn to the $j = 1$ case, which is

$$\mathbb{E}\sum_{n\in\mathbf{S}\cap[1,y]}\sum_{q\in\mathcal{Q}_H}\sum_{h\leqslant KH}\boldsymbol{\lambda}(H;q,n-qh) = \Big(1+O\Big(\frac{1}{H^{M-2}}\Big)\Big)|\mathcal{Q}_H|KH\sigma_1 y.$$

It suffices to show that for each $h \leqslant KH$, one has

$$\mathbb{E}\sum_{n\in\mathbf{S}\cap[1,y]}\sum_{q\in\mathcal{Q}_H}\boldsymbol{\lambda}(H;q,n-qh) = \Big(1+O\Big(\frac{1}{H^{M-2}}\Big)\Big)|\mathcal{Q}_H|\sigma_1 y. \tag{5.9}$$

The left-hand side can be expanded as

$$\mathbb{E} \sum_{n\in \mathbf{S}\cap[1,y]} \sum_{q\in\mathcal{Q}_H} \frac{1_{\mathbf{AP}(KH;q,n-qh)\subset \mathbf{S}_2}}{\sigma_2^{|\mathbf{AP}(KH;q,n-qh)|}}.$$

By (2.11), the constraint $n \in \mathbf{S}\cap[1,y]$ implies that $n \in \mathbf{S}_1\cap[1,y]$. Conversely, if $n \in \mathbf{S}_1\cap[1,y]$, then $n \in \mathbf{AP}(H; q, n-qh)$, and the condition $n \in \mathbf{S}$ is subsumed in the condition that $\mathbf{AP}(KH; q, n-qh) \subset \mathbf{S}_2$. Thus we may replace the constraint $n \in \mathbf{S} \cap [1,y]$ here with $n \in \mathbf{S}_1 \cap [1,y]$ and rewrite the above expression as

$$\mathbb{E} \sum_{n\in \mathbf{S}_1\cap[1,y]} \sum_{q\in\mathcal{Q}_H} \frac{1_{\mathbf{AP}(KH;q,n-qh)\subset \mathbf{S}_2}}{\sigma_2^{|\mathbf{AP}(KH;q,n-qh)|}}.$$

Recall that $\mathbf{S}_2$ is independent of $\mathbf{S}_1$ and of $\mathbf{AP}(KH; q, n-qh)$. Applying Lemma 5.1 as before, we may write the left side of (5.9) as

$$\mathbb{E} \sum_{n\in \mathbf{S}_1\cap[1,y]} \sum_{q\in\mathcal{Q}_H} \left(1 + O\left(\frac{1}{H^{M-2}}\right) + O\left(\frac{1}{H^2} \sum_{\substack{h',h''\leqslant KH\\ h'\neq h''}} E_{8BK^2H^2}(qh'-qh'')\right)\right).$$

Trivially we have

$$\mathbb{E}|\mathbf{S}_1\cap[1,y]| = \sum_{n=1}^{y} \mathbb{P}(n\in\mathbf{S}_1) = \sigma_1 y, \tag{5.10}$$

and the claim (5.9) now follows from (5.8).

Finally, we establish the $j=2$ case of Theorem 3(iii), which expands as

$$\sum_{h_1,h_2\leqslant KH} \mathbb{E} \sum_{n\in\mathbf{S}\cap[1,y]} \sum_{q_1,q_2\in\mathcal{Q}_H} \boldsymbol{\lambda}(H;q_1,n-q_1h_1)\boldsymbol{\lambda}(H;q_2,n-q_2h_2) = \\ = \left(1+O\left(\frac{1}{H^{M-2}}\right)\right)|\mathcal{Q}_H|^2K^2H^2\frac{\sigma_1}{\sigma_2}y.$$

With h_1, h_2 fixed, we can use (2.14) to expand the sum over n, q_1, q_2 as

$$\mathbb{E} \sum_{n\in\mathbf{S}\cap[1,y]} \sum_{q_1,q_2\in\mathcal{Q}_H} \frac{1_{\mathbf{AP}(KH;q_1,n-q_1h_1)\cup\mathbf{AP}(KH;q_2,n-q_2h_2)\subset\mathbf{S}_2}}{\sigma_2^{|\mathbf{AP}(KH;q_1,n-q_1h_1)|+|\mathbf{AP}(KH;q_2,n-q_2h_2)|}} \tag{5.11}$$

As in the $j=1$ case, we may replace the constraint $n \in \mathbf{S}\cap[1,y]$ here with $n\in\mathbf{S}_1\cap[1,y]$. Next, we observe that the set

$$\mathbf{AP}(KH;q_1,n-q_1h_1)\cup\mathbf{AP}(KH;q_2,n-q_2h_2)$$

contains at most $|\mathbf{AP}(KH;q_1,n-q_1h_1)| + |\mathbf{AP}(KH;q_2,n-q_2h_2)| - 1$ distinct elements, as n is common to both of the sets $\mathbf{AP}(KH;q_1,n-q_1h_1), \mathbf{AP}(KH;q_2,n-q_2h_2)$. Thus if we apply Lemma 5.1 (noting that $\mathbf{S}_2$ is independent of $\mathbf{S}_1, \mathbf{AP}(KH;q_1,n-q_1h_1)$ and $\mathbf{AP}(KH;q_2,n-q_2h_2)$) after eliminating the duplicate constraint, we may write (5.11) as

$$\sigma_2^{-1}\mathbb{E} \sum_{n\in\mathbf{S}_1\cap[1,y]} \sum_{q_1,q_2\in\mathcal{Q}_H} \left(1+O\left(\frac{1}{H^{M-2}} + \frac{E'(q_1)+E'(q_2)+E''(q_1,q_2)}{H^2}\right)\right)$$

where

$$E'(q) := \sum_{\substack{h,h' \leqslant KH \\ h \neq h'}} E_{8BK^2H^2}(qh - qh'; H)$$

and

$$E''(q_1, q_2) := \sum_{\substack{h_1', h_2' \leqslant KH \\ h_1 \neq h_1', h_2 \neq h_2'}} E_{8BK^2H^2}(q_1 h_1' - q_1 h_1 - q_2 h_2' + q_2 h_2; H).$$

The average over $E'(q_1) + E'(q_2)$ is acceptably small by the $j = 1$ analysis. Thus (using (5.10)) it suffices to show that

$$\sum_{q_1, q_2 \in \mathcal{Q}_H} E_{8BK^2H^2}(q_1 h_1' - q_1 h_1 - q_2 h_2' + q_2 h_2; H) \ll \frac{1}{H^{M-2}} |\mathcal{Q}_H|^2$$

for each $h_1', h_2' \leqslant KH$ with $h_1' \neq h_1$, $h_2' \neq h_2$). But this follows from (5.8) (applied with q replaced by q_1 and k replaced by $-q_2 h_2' + q_2 h_2$, and then summing in q_2). This completes the proof of the $j = 2$ case, and so establishes (4.4). □

We have now verified all the the claims (4.1)-(4.4), and so have completed the proof of Theorem 3.

Appendix A. Proof of the covering lemma

In this appendix we prove Lemma 3.1. Our main tool will be the following general hypergraph covering lemma from [5, Theorem 3]:

Theorem A (Probabilistic covering)**.** *There exists an absolute constant $C_4 \geqslant 1$ such that the following holds. Let $D, r, A \geqslant 1$, $0 < \kappa \leqslant 1/2$, and let $m \geqslant 0$ be an integer. Let $\tau > 0$ satisfy*

$$\tau \leqslant \left(\frac{\kappa^A}{C_4 \exp(AD)} \right)^{10^{m+2}}. \tag{A.1}$$

Let $I_1, \dots, I_m$ be disjoint finite non-empty sets, and let V be a finite set. For each $1 \leqslant j \leqslant m$ and $i \in I_j$, let $\mathbf{e}_i$ be a random subset of V. Assume the following:

- *(Edges not too large) Almost surely for all $j = 1, \dots, m$ and $i \in I_j$, we have*

$$\#\mathbf{e}_i \leqslant r; \tag{A.2}$$

- *(Each sieve step is sparse) For all $j = 1, \dots, m$, $i \in I_j$ and $v \in V$,*

$$\mathbb{P}(v \in \mathbf{e}_i) \leqslant \frac{\tau}{|I_j|^{1/2}}; \tag{A.3}$$

- *(Very small codegrees) For every $j = 1, \dots, m$, and distinct $v_1, v_2 \in V$,*

$$\sum_{i \in I_j} \mathbb{P}(v_1, v_2 \in \mathbf{e}_i) \leqslant \tau \tag{A.4}$$

- *(Degree bound) If for every $v \in V$ and $j = 1, \dots, m$ we introduce the normalized degrees*

$$d_{I_j}(v) := \sum_{i \in I_j} \mathbb{P}(v \in \mathbf{e}_i) \tag{A.5}$$

and then recursively define the quantities $P_j(v)$ for $j = 0, \dots, m$ and $v \in V$ by setting

$$P_0(v) := 1 \tag{A.6}$$

and

$$P_{j+1}(v) := P_j(v) \exp(-d_{I_{j+1}}(v)/P_j(v)) \tag{A.7}$$

for $j = 0, \dots, m-1$ and $v \in V$, then we have

$$d_{I_j}(v) \leqslant DP_{j-1}(v) \qquad (1 \leqslant j \leqslant m, v \in V)$$

and

$$P_j(v) \geqslant \kappa \qquad (0 \leqslant j \leqslant m, v \in V).$$

Then there are random variables $\mathbf{e}'_i$ for each $i \in \bigcup_{j=1}^m I_j$ with the following properties:

(a) *For each $i \in \bigcup_{j=1}^m I_j$, the support of $\mathbf{e}'_i$ is contained in the support of $\mathbf{e}_i$, union the empty set singleton $\{\emptyset\}$. In other words, almost surely $\mathbf{e}'_i$ is either empty, or is a set that $\mathbf{e}_i$ also attains with positive probability.*

(b) *For any $0 \leqslant J \leqslant m$ and any finite subset e of V with $\#e \leqslant A - 2rJ$, one has*

$$\mathbb{P}\left(e \subset V \backslash \bigcup_{j=1}^{J} \bigcup_{i \in I_j} \mathbf{e}'_i\right) = \left(1 + O(\tau^{1/10^{J+1}})\right) P_J(e)$$

where

$$P_j(e) := \prod_{v \in e} P_j(v).$$

Proof. See [5, Theorem 3]. □

To derive Lemma 3.1 from Theorem A, we repeat the proof of [5, Corollary 4] with a different choice of parameters. Let the notation and hypotheses be as in Lemma 3.1. Firstly, we may assume that $\eta \leqslant \frac{1}{1000}$, for the conclusion is trivial otherwise.

Let $\beta = \beta(\delta)$ be a parameter satisfying

$$\beta > 10^{2\delta} > \frac{\beta \log \beta}{\beta - 1} \tag{A.8}$$

This is possible as $\log \beta < \beta - 1$ for all $\beta > 1$. Let

$$m = \left\lceil \frac{\log(1/\eta)}{\log \beta} \right\rceil \tag{A.9}$$

so that, by (3.9),

$$1 \leqslant m \leqslant \frac{\delta \log\log y + \log\log\log y}{\log \beta} + 1, \qquad \frac{1}{\eta} \leqslant \beta^m \leqslant \frac{\beta}{\eta}. \tag{A.10}$$

By (3.9) and (A.8), $C_2 > \frac{\beta \log \beta}{\beta - 1}$ and thus we may find disjoint intervals $\mathscr{I}_1, \dots, \mathscr{I}_m$ in $[0,1]$ with length

$$|\mathscr{I}_j| = \frac{\beta^{1-j} \log \beta}{C_2} \qquad (1 \leqslant j \leqslant m). \tag{A.11}$$

Let $\vec{\mathbf{t}} = (\mathbf{t}_1, \dots, \mathbf{t}_s)$, where $\mathbf{t}_i$ is a uniform random real number in $[0,1]$ for each i, and such that $\mathbf{t}_1, \dots, \mathbf{t}_s$ are independent. Define the random sets

$$I_j = I_j(\vec{\mathbf{t}}) := \{1 \leqslant i \leqslant s : \mathbf{t}_i \in \mathscr{I}_j\}$$

for $j = 1, \dots, m$. These sets are clearly disjoint.

We will verify (for a suitable choice of $\vec{\mathbf{t}}$) the hypotheses of Theorem A with the indicated sets I_j and random variables $\mathbf{e}_i$, and with suitable choices of parameters $D, r, A \geqslant 1$ and $0 < \kappa \leqslant 1/2$.

Let $v \in V$, $1 \leqslant j \leqslant m$ and consider the independent random variables $(\mathbf{X}_i^{(v,j)}(\vec{\mathbf{t}}))_{1 \leqslant i \leqslant s}$, where

$$\mathbf{X}_i^{(v,j)}(\vec{\mathbf{t}}) = \begin{cases} \mathbb{P}(v \in \mathbf{e}_i) & \text{if } i \in I_j(\vec{\mathbf{t}}) \\ 0 & \text{otherwise.} \end{cases}$$

By (3.8), (A.11), and (A.10), we have for every $1 \leqslant j \leqslant m$ and $v \in V$ that

$$\begin{aligned} \sum_{i=1}^{s} \mathbb{E}\mathbf{X}_i^{(v,j)}(\vec{\mathbf{t}}) &= \sum_{i=1}^{s} \mathbb{P}(v \in \mathbf{e}_i)\mathbb{P}(i \in I_j(\vec{\mathbf{t}})) \\ &= |\mathscr{I}_j| \sum_{i=1}^{s} \mathbb{P}(v \in \mathbf{e}_i) \\ &= \beta^{1-j} \log \beta + O\big(\eta \beta^{-j} \log \beta\big) \\ &= \beta^{1-j} \log \beta + O\big(\beta^{-m-j} \log \beta\big). \end{aligned}$$

In the last equality we have used that $C_2 \geqslant 1$.

By (3.6), we have $|\mathbf{X}_i^{(v,j)}(\vec{\mathbf{t}})| \leqslant y^{-1/2-1/100}$ for all i, and hence by Hoeffding's inequality,

$$\begin{aligned} \mathbb{P}\left(\Big| \sum_{i=1}^{s} (\mathbf{X}_i^{(v,j)}(\vec{\mathbf{t}}) - \mathbb{E}\mathbf{X}_i^{(v,j)}(\vec{\mathbf{t}})) \Big| \geqslant \frac{1}{y^{1/200}} \right) &\leqslant 2 \exp\left\{ -2 \frac{y^{-1/100}}{y^{-1-1/50} s} \right\} \\ &= 2 \exp\left\{ -2y^{1/100} \right\}. \end{aligned}$$

Here we used the hypothesis $s \leqslant y$. By a union bound, the bound $|V| \leqslant y$ and (A.9), there is a deterministic choice $\vec{t}$ of $\vec{\mathbf{t}}$ (and hence $I_1, \dots, I_m$) such that for *every* $v \in V$ and *every* $j = 1, \dots, m$, we have

$$\Big| \sum_{i=1}^{s} (\mathbf{X}_i^{(v,j)}(\vec{t}) - \mathbb{E}\mathbf{X}_i^{(v,j)}(\vec{\mathbf{t}})) \Big| < \frac{1}{y^{1/200}}.$$

Note that this is vastly smaller than $\beta^{-m} \asymp (\log y)^{-\delta}$. We fix this choice $\vec{t}$ (so that the I_j are now deterministic), and we conclude that for y sufficiently large (in terms of δ)

$$\begin{aligned}\sum_{i\in I_j} \mathbb{P}(v \in \mathbf{e}_i) &= \sum_{i=1}^{s} \mathbf{X}_i^{(v,j)}(\vec{t}) \\ &= \beta^{1-j}\log\beta + O\left(\beta^{-j-m}\log\beta + \frac{1}{y^{1/200}}\right) \\ &= \beta^{1-j}\log\beta + O\left(\beta^{-j-m}\log\beta\right)\end{aligned} \tag{A.12}$$

uniformly for all $j = 1, \dots, m$, and all $v \in V$. In particular, all sets I_j are nonempty.

Set

$$\tau := y^{-1/100} \tag{A.13}$$

and observe from (3.6) and the bound $|I_j| \leqslant s \leqslant y$ that the sparsity condition (A.3) holds. Also, the small codegree condition (3.7) implies the small codegree condition (A.4).

From (A.5), (A.12) and (A.10), we now have

$$d_{I_j}(v) = (1 + O(\beta^{-m}))\beta^{-j+1}\log\beta$$

for all $v \in V$, $1 \leqslant j \leqslant m$. Let λ satisfy $1 + \log\beta < \lambda < \beta$. A routine induction using (A.6), (A.7) then shows (for y sufficiently large) that

$$P_j(v) = (1 + O(\lambda^j\beta^{-m}))\beta^{-j} \quad (0 \leqslant j \leqslant m), \tag{A.14}$$

In particular we have

$$d_{I_j}(v) \leqslant DP_{j-1}(v) \qquad (1 \leqslant j \leqslant m)$$

for some absolute constant D, and

$$P_j(v) \geqslant \kappa \qquad (0 \leqslant j \leqslant m),$$

where

$$\kappa \gg \beta^{-m} \geqslant \eta/\beta \gg \eta.$$

We now set

$$r = \frac{(\log y)^{1/2}}{\log\log y}, \qquad A := 2rm + 1.$$

By (A.10) and (3.5), one has

$$A \ll (\log y)^{1/2}$$

and so (A.2) holds and also

$$\frac{\kappa^A}{C_4\exp(AD)} \gg \exp\left(-O\left((\log y)^{1/2}(\log\log y)\right)\right). \tag{A.15}$$

By (A.9) and (A.8),

$$10^m \ll (1/\eta)^{\frac{\log 10}{\log\beta}} \ll (\log y)^{\frac{\delta\log 10}{\log\beta}}(\log\log y)^{\frac{\log 10}{\log\beta}} < (\log y)^{1/2-\varepsilon_1}$$

for some $\varepsilon_1 = \varepsilon_1(\delta) > 0$. Hence by (A.13), we see that

$$\tau^{1/10^{m+2}} \leqslant \exp\left\{-K(\log y)^{1/2+\varepsilon_1}\right\}, \tag{A.16}$$

for some absolute constant $K > 0$. Combining (A.15) and (A.16), we see that (A.1) is satisfied if y is large enough. Thus all the hypotheses of Theorem A have been verified for this choice of parameters. Applying this Theorem A and using (A.14), one thus obtains random variables $\mathbf{e}_i'$ for $i \in \bigcup_{j=1}^m I_j$ whose range is contained in the range of $\mathbf{e}_i$ together with $\emptyset$, such that

$$\mathbb{P}\left(n \notin \bigcup_{j=1}^{m} \bigcup_{i \in I_j} \mathbf{e}_i'\right) \ll \beta^{-m} \ll \eta$$

for all $n \in V$. For $1 \leqslant i \leqslant s$, $i \notin \bigcup_{j=1}^m I_j$, set $\mathbf{e}_i' = \emptyset$ with probability 1. By linearity of expectation this gives

$$\mathbb{E}\Big|V \setminus \bigcup_{i=1}^{s} \mathbf{e}_i'\Big| \ll \eta |V|.$$

Hence, for some absolute constant $C_3 > 0$, we have

$$\Big|V \setminus \bigcup_{i=1}^{s} \mathbf{e}_i'\Big| \leqslant C_3 \eta |V|$$

with probability $\geqslant 1/2$. Therefore, there is some vector $(e_1, \dots, e_s)$ of subsets of V, where, for every i, e_i is in the support of $\mathbf{e}_i$ or is the empty set, for which (3.10) holds. Finally, for the i such that e_i is the empty set, replace e_i with an arbitrary element in the support of $\mathbf{e}_i$; clearly (3.10) still holds.

REFERENCES

[1] P. T. Bateman and R. A. Horn, *A heuristic asymptotic formula concerning the distribution of prime numbers*, Math. Comp. **16** (1962), 363–367.

[2] V. Bouniakowsky, *Nouveaux théorèmes relatifs á la distinction des nombres premiers et á la d'ecomposition des entiers en facteurs*, Mém. Acad. Sc. St. Pétersbourg **6** (1857), 305–329.

[3] N. Tschebotareff, *Die Bestimmung der Dichtigkeit einer Menge von Primzahlen, welche zu einer gegebenen Substitutionsklasse gehören*, Mathematische Annalen **95** (1) (1926), 191–228.

[4] A. C. Cojocaru and M. R. Murty, *An introduction to Sieve Methods and their Applications*, Cambridge University Press, 2006.

[5] K. Ford. B. Green, S. Konyagin, J. Maynard, and T. Tao, *Long gaps between primes*, J. Amer. Math. Soc. **31** (2018), no. 1, 65–105.

[6] J. Friedlander and H. Iwaniec, *Opera de Cribro*, Amer. Math. Soc., 2010.

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

[8] C. Hooley, *Applications of sieve methods to the theory of numbers*, Cambridge Tracts in Mathematics, No. 70, Cambridge University Press, 1976.

[9] J. C. Lagarias and A. M. Odlyzko, *Effective versions of the Chebotarev density theorem*, Algebraic number fields: L-functions and Galois properties (Proc. Sympos., Univ. Durham, Durham, 1975), Academic Press, 1977, pp. 409–464.

[10] E. Landau, *Neuer Beweis des Primzahlsatzes und Beweis des Primidealsatzes*, Mathematische Annalen. **56**, No. 4, (1903), 645–670.

[11] G. Pólya, *Über ganzwertige ganze Funktionen*, Rend. Circ. Mat. Palermo **40** (1915), 1–16.

[12] C. Sanna and M. Szikszai, *A coprimality condition on consecutive values of polynomials*, Bull. London Math. Soc. (2017), DOI: 10.1112/blms.12078. Also see arXiv:1704.01738v1.

[13] B. L. van der Waerden. *Die Seltenheit der reduziblen Gleichungen und der Gleichungen mit Affekt.*, Monatsh. Math. Phys. **43**(1) (1936), 133–147.

(Corresponding author) DEPARTMENT OF MATHEMATICS, 1409 WEST GREEN STREET, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, URBANA, IL 61801, USA
E-mail address: ford126@illinois.edu

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

MATHEMATICAL INSTITUTE, RADCLIFFE OBSERVATORY QUARTER, WOODSTOCK ROAD, OXFORD OX2 6GG, ENGLAND
E-mail address: james.alexander.maynard@gmail.com

MATHEMATICS DEPARTMENT, DARTMOUTH COLLEGE, HANOVER, NH 03755, USA
E-mail address: carl.pomerance@dartmouth.edu

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