Additive decompositions of sets with restricted prime factors

Christian Elsholtz, Adam J. Harper*

January 30, 2014

Abstract

We investigate sumset decompositions of quite general sets with restricted prime factors. We manage to handle certain sets, such as the smooth numbers, even though they have little sieve amenability, and conclude that these sets cannot be written as a ternary sumset. This proves a conjecture by Sárközy. We also clean up and sharpen existing results on sumset decompositions of the prime numbers.

1 Introduction

At the beginning of his two volume treatise on additive number theory [27] Ostmann introduces the concept of asymptotic additive decomposability of a set:

Definition 1.1 (See [27], vol. 1, p. 5)**.** *Let $\mathcal{S}$ be a set of positive integers. We say that $\mathcal{S}$ is* asymptotically additively decomposable (or reducible) *if there are two sets of positive integers $\mathcal{A}$ and $\mathcal{B}$, with at least two elements each, such that for a sufficiently large x_0:*

$$(\mathcal{A} + \mathcal{B}) \cap [x_0, \infty) = \mathcal{S} \cap [x_0, \infty).$$

Here the sumset of $\mathcal{A}$ and $\mathcal{B}$ is defined by $\mathcal{A} + \mathcal{B} := \{a + b : a \in \mathcal{A}, b \in \mathcal{B}\}$.

If $\mathcal{S}$ is asymptotically additively decomposable we write this as $\mathcal{S} \sim \mathcal{A} + \mathcal{B}$. Otherwise we say that $\mathcal{S}$ is asymptotically additively indecomposable *(or* irreducible*).*

Wirsing [39] showed that almost all sets of integers are asymptotically additively irreducible. But it seems very difficult to prove whether a given set $\mathcal{S}$ that "occurs in nature" is asymptotically additively irreducible or not. Shortly we will describe some of the results obtained previously. The main purpose of this paper is to develop a more structural approach, rather than only looking at special examples, and to identify a large family of multiplicatively defined sets (including sets of smooth numbers, the set of primes, and sets of integers composed from dense subsequences of the primes) for which one has the same consequences: if the sets can be asymptotically additively decomposed, then both summands must have counting functions about $x^{1/2+o(1)}$ (see Theorem 4.1 for a general statement of this type); moreover there is no asymptotic additive decomposition into three sets.

Methodologically, the new ingredients of this paper are the use of Selberg's sieve, in combination with the large and larger sieves, to handle less regular sets; and the use of estimates for sums of general multiplicative functions, to make the process of combining the sieve arguments more efficient and general. The former is ultimately what allows us to treat the smooth numbers example, and the latter is what allows us to abstract the arguments into the general Theorem 4.1, and to obtain some improvements in the prime numbers example. We also borrow an idea of Sárközy [34] from the finite field case, by using a sumset inequality of Ruzsa [30] to deduce ternary indecomposability results from our binary counting results. This is simpler and more widely applicable than the previous local approach to such deductions.

The most frequently studied case in the literature is Ostmann's conjecture.

Conjecture 1.2 (Ostmann, [27], volume 1, page 13)**.** *The set $\mathcal{P}$ of primes is asymptotically additively irreducible.*

This conjecture, which is sometimes called the "inverse Goldbach conjecture", has sparked considerable interest. Erdős [10] called it "an old and very fascinating conjecture", and in [11] he called it "one of the most interesting unconventional problems on primes", and he included it in several more problem collections. A variety of methods have been applied to try to attack the problem: congruence, gap or density considerations (Ostmann [27], Laffer and Mann [23], Hornfeck (see Ostmann [27], volume I, page 13), Puchta [29]); the small sieve (Hornfeck [22], Mann [24], Croot and Elsholtz [2]); the large sieve (Pomerance, Sárközy and Stewart [28], Hofmann and Wolke [21], Elsholtz [4], Croot and Elsholtz [2]); and a combination of the large sieve method with Gallagher's larger sieve (Elsholtz [5], [8]). See the first author's paper [8] for some further historical references.

It is known that a ternary decomposition $\mathcal{P} \sim \mathcal{A} + \mathcal{B} + \mathcal{C}$ is impossible, where $\mathcal{A}, \mathcal{B}, \mathcal{C}$ are sets of positive integers with at least two elements each

(see [5], [6]). It is also known that if $\mathcal{P} \sim \mathcal{A} + \mathcal{B}$, then $\max(\mathcal{A}(x), \mathcal{B}(x)) \ll x^{1/2}(\log x)^2$ (see [8]) and $\frac{x}{\log x} \ll \mathcal{A}(x)\mathcal{B}(x) \ll x$ (see [41], [28], and a survey [8]). Here $\mathcal{A}(x) := \#\{n \le x : n \in \mathcal{A}\}$ is the counting function, similarly for the counting function $\mathcal{B}(x)$.

Some other related problems have been studied. For example, it is known that the set of shifted primes $\mathcal{P} + c$ cannot be asymptotically decomposed into a product set $\mathcal{AB}$, see [9].

Other sets of integers have been investigated for decomposability but Erdős and Nathanson [12] write "very little is understood about the general problem of the decomposition of sets in the form $S = A + B$, where $|A| \ge 2$ and $|B| \ge 2$." For some general results on additive decomposability of integers see Sárközy [31, 32]. The set of integer squares is irreducible, simply because the gaps between squares are increasing, but Erdős conjectured that the set of squares $\{n^2 \le N^2\}$ remains irreducible even if one is allowed to change $o(N)$ of the elements. Partial results are due to Sárközy and Szemerédi [35].

We also mention that various work has been done recently on finite field versions of the above questions. To give a few examples, Sárközy [34] investigated the decomposability of the set of squares modulo primes, and Dartyge and Sárközy [3] investigated the set of primitive roots modulo primes. Extensions of their results in various directions are due to Shparlinski [37], Shkredov [36], and Gyarmati, Konyagin and Sárközy [18]. One also has finite field analogues of Wirsing's result [39] mentioned above, for example Alon, Granville and Ubis [1] proved that the number of sets $\mathcal{S} \subset \mathbb{Z}/p\mathbb{Z}$ that can be written as a sumset $\mathcal{S} = \mathcal{A} + \mathcal{B}$ is small, namely $2^{p/2+o(p)}$.

Following the solution [5] of the ternary version of Ostmann's problem, mentioned above, Sárközy [33] posed the binary and ternary additive decomposability of the smooth numbers as an open problem. A transfer of the methods from [5] did not appear to work.

Definition 1.3. *A number is said to be y-smooth if all of its prime factors are at most y. We will write $\mathcal{S}_y$ for the set of all y-smooth numbers. More generally, if $f(n)$ is a function, then we will write $\mathcal{S}_{f(n)}$ for the set of all numbers n that are $f(n)$-smooth.*

Conjecture 1.4 (Sárközy [33])**.** *Let $0 < \epsilon < 1$ be fixed. Then the set $\mathcal{S}_{n^\epsilon}$, of integers n with no prime factor greater than n^ϵ, is asymptotically additively irreducible.*

2 Announcement of results

2.1 Sets of smooth numbers

An argument along the lines of [5] does not seem to work unless the size $\mathcal{S}(x)$ of the target set can be well estimated by a sieving procedure. This excludes some natural sets, such as sets $\mathcal{S}_{f(n)}$ of smooth numbers, from consideration. In this paper we will prove an additive irreducibility theorem for sets that need not be well controlled by the sieve in this sense, but have certain other regularity properties. We defer a general statement, which is the main result of this paper but is a bit technical, to Theorem 4.1, but as a special case we achieve the following results on smooth numbers. In particular, the ternary version of the above mentioned conjecture of Sárközy will be solved for all small ϵ.

Theorem 2.1. *There exist a large absolute constant $D > 0$, and a small absolute constant $\kappa > 0$, such that the following is true. Suppose $f(n)$ is an increasing function such that $\log^D n \leq f(n) \leq n^\kappa$ for large n, and such that $f(2n) \leq f(n)(1 + (100 \log f(n))/\log n)$. Then if $\mathcal{A} + \mathcal{B} \sim \mathcal{S}_{f(n)}$, where $\mathcal{A}$ and $\mathcal{B}$ contain at least two elements each, we have*

$$\max(\mathcal{A}(x), \mathcal{B}(x)) \ll \sqrt{x} \log^4 x.$$

Corollary 2.2. *If $f(n)$ is as in Theorem 2.1, then a ternary decomposition $\mathcal{A} + \mathcal{B} + \mathcal{C} \sim \mathcal{S}_{f(n)}$, where $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$ contain at least two elements each, does not exist.*

In particular, one can take $f(n) = n^\epsilon$ in Corollary 2.2 for any fixed $0 < \epsilon \leq \kappa$.

Let us explain briefly how we are able to handle the smooth numbers case, referring the reader to section 4 for fuller details. The proof strategy of [5], where the primes were handled, can be summarized as follows: suppose that $\mathcal{A} + \mathcal{B} \sim \mathcal{S}$, where the set $\mathcal{S}$ shall be investigated. Since $\mathcal{S}$ has restricted prime factors, a sieve argument shows that if $\mathcal{A}(x)$ is a certain size then $\mathcal{B}(x)$ is rather less than $\frac{\mathcal{S}(x)}{\mathcal{A}(x)}$, which is a contradiction since we certainly have

$$\mathcal{S}(x) + O(1) = (\mathcal{A} + \mathcal{B})(x) \leq \mathcal{A}(x)\mathcal{B}(x)$$

if $\mathcal{A} + \mathcal{B} \sim \mathcal{S}$. Studying different ranges of $\mathcal{A}(x)$ this eventually leads to good lower and upper bounds on $\mathcal{A}(x)$ and $\mathcal{B}(x)$. Then one can deduce a ternary indecomposability result, which in this paper will be done using general sumset inequalities.

If $\mathcal{S}$ is a set of smooth numbers, and if one tries a direct application of the sieve methods applied to the set of primes, then one gets stuck right at the beginning when trying to find basic upper bounds for k-tuples of smooth

numbers. For with $u = \frac{\log x}{\log y}$, and ϱ denoting the Dickman function, it is known that $\mathcal{S}_y(x) \sim \varrho(u)x \approx \frac{x}{u^u}$ on a wide range. If $a_1, \ldots, a_k$ are distinct integers one would therefore expect that

$$\#\{n \leq x : n + a_1, \ldots, n + a_k \in \mathcal{S}_y\} \approx \varrho(u)^k x \approx \frac{x}{u^{uk}}.$$

By a usual application of sieve methods (sieving the integers $n \leq x$ by primes $y < p \leq \sqrt{x}$, say) one would rather prove something like

$$\#\{n \leq x : n + a_1, \ldots, n + a_k \in \mathcal{S}_y\} \ll_k \frac{x}{u^k}.$$

If $k < u$ this estimate is even worse than the "trivial" bound $\mathcal{S}_y(x)$ for the left hand side, and so the method outlined above cannot proceed. Note that in the primes case one does obtain correct order upper bounds for prime k-tuples.

In contrast to this, we will use Selberg's sieve and some other ingredients (see the proof of Proposition 4.3, below) to show something like

$$\#\{n \leq x : n + a_1, \ldots, n + a_k \in \mathcal{S}_y\} \ll_k \frac{x}{u^{u+(k-1)}},$$

and will show that this is sufficient to make our proof run.

2.2 Sets consisting of prime factors from a certain set

As a second example we study semigroup-type sets $\mathcal{Q}(\mathcal{T})$ of integers composed of certain prime factors $\mathcal{T}$ only.

Theorem 2.3. *Let $\mathcal{T}$ be any finite set of primes, and let*

$$\mathcal{Q}(\mathcal{T}) = \{n \in \mathbb{N} : p | n \Rightarrow p \in \mathcal{T}\}.$$

Then $\mathcal{Q}(\mathcal{T})$ does not have an asymptotic additive decomposition into two sets.

Proof. Improving results of Thue, Siegel, Mahler and making use of Baker's theory on linear forms in logarithms, Tijdeman [38, page 319] proved that the sequence (q_n) of those integers composed of finitely many given primes has very large gaps. Here we only make use of the weaker statement that $\liminf_{n \to \infty} (q_{n+1} - q_n) = \infty$. On the other hand, if $\mathcal{A} + \mathcal{B} \sim \mathcal{Q}(\mathcal{T})$ and $a_1, a_2 \in \mathcal{A}$, and $\mathcal{B}$ is infinite, then $a_2 - a_1$ must be a (bounded) gap of infinitely many pairs of elements of $\mathcal{Q}(\mathcal{T})$. This is a contradiction. $\qquad \square$

From Tijdeman [38, Theorem 7] it similarly follows that there exists an infinite set $\mathcal{T}$ with the same conclusion.

Using the sieve-based methods of this paper, rather than gap considerations, we shall prove some related results.

Theorem 2.4. *Let $\mathcal{T}$ denote any set of primes with*

$$\sum_{\substack{p \leq x \\ p \in \mathcal{T}}} \frac{\log p}{p} = \tau \log x + C + o(1).$$

Here $0 < \tau < 1$ and C denote any real constants. Let

$$\mathcal{Q}(\mathcal{T}) = \{n \in \mathbb{N} : p|n \Rightarrow p \in \mathcal{T}\}.$$

If $\mathcal{Q}(\mathcal{T}) \sim \mathcal{A} + \mathcal{B}$, where $\mathcal{A}, \mathcal{B}$ contain at least two elements each, then

$$\max(\mathcal{A}(x), \mathcal{B}(x)) \ll_{\mathcal{T}} x^{1/2} (\log x)^4.$$

This theorem includes the case where $\mathcal{T}$ consists of the primes in one or several arithmetic progressions.

Corollary 2.5. *Let $\mathcal{T}$ satisfy the hypotheses of Theorem 2.4. Then the set $\mathcal{Q}(\mathcal{T}) = \{n \in \mathbb{N} : p|n \Rightarrow p \in \mathcal{T}\}$ cannot be asymptotically decomposed into three sets with at least two elements each.*

It was previously proved in [8] that the product

$$\mathcal{A}(x)\mathcal{B}(x) \ll_{\tau,C} x(\log x)^{2\tau},$$

and that there are various examples, based on properties of sums of two squares (which is a class of examples of some conjectural relevance to Ostmann's problem, see section 2.3), where $\mathcal{A} + \mathcal{B}$ is a large subset of $\mathcal{Q}(\mathcal{T})$ and

$$\mathcal{A}(x)\mathcal{B}(x) \gg_{\tau,C} x(\log x)^{-1/2}.$$

Obtaining a bound on $\max(\mathcal{A}(x), \mathcal{B}(x))$, as we do in this paper as a straightforward corollary of our general Theorem 4.1, is a stronger kind of information and allows us to deduce Corollary 2.5.

2.3 Ostmann's problem revisited

Theorem 2.6. *Suppose that $\mathcal{P} \sim \mathcal{A} + \mathcal{B}$, where $\mathcal{P}$ is the set of all primes and $\mathcal{A}, \mathcal{B}$ contain at least two elements each. Then the following bounds hold:*

$$\frac{x^{1/2}}{\log x \log \log x} \ll \mathcal{A}(x) \ll x^{1/2} \log \log x.$$

The same bounds hold for $\mathcal{B}(x)$.

It was known before that $\mathcal{A}(x)\mathcal{B}(x) \ll x$ (for a survey of references see [8]) and $\max(\mathcal{A}(x), \mathcal{B}(x)) \ll x^{1/2}(\log x)^2$ in the infinite version of the

problem[1]. In this case, our improvement is derived by using a lower bound estimate of Hildebrand [20] for sums of multiplicative functions to make part of the sieving argument more efficient. See section 7, below.

It is an odd feature of Ostmann's problem that we know that modulo many primes, the putative summand sets $\mathcal{A}$ and $\mathcal{B}$ both must lie in about $\frac{p}{2}$ residue classes. See [5], or the proof of Theorem 2.6 in section 7. Heuristically one would therefore expect that the sets $\mathcal{A}$ and $\mathcal{B}$ must have some special structure, i.e. are essentially the values of a quadratic polynomial. So $\mathcal{A} + \mathcal{B}$ must resemble somewhat a binary quadratic form, which we certainly wouldn't expect to coincide with the set of primes. But currently this heuristic cannot be substantiated rigorously. Some results in this direction are obtained in the paper [17] of Green and Harper.

3 Some background results

3.1 Lemmas from sieve theory

The main ingredients of our proofs are Selberg's sieve, Montgomery's large sieve inequality, and Gallagher's larger sieve, combined in various ways. We state versions of these sieve results now.

Lemma 3.1 (Selberg, see Theorem 7.1 of [14]). *Let $\mathcal{P}_0$ denote a finite subset of the primes, and let $\mathcal{C}$ be a finite, non-empty set of integers. Let $r_1, \ldots, r_k$ be distinct integers. Define a set $\mathcal{B} \subseteq \mathcal{C}$ in the following way:*

$$\mathcal{B} := \{c \in \mathcal{C} : c \not\equiv r_i \bmod p, \ 1 \leq i \leq k, \ p \in \mathcal{P}_0\}.$$

Finally, suppose that $\omega_{\mathcal{B}} : \mathcal{P}_0 \to \mathbb{R}$ is any function with $0 \leq \omega_{\mathcal{B}}(p) < p$, and that Q is any positive integer. Then we have

$$\#\mathcal{B} \leq \frac{\#\mathcal{C}}{L} + \sum_{\substack{d \leq Q^2, \\ p|d \Rightarrow p \in \mathcal{P}_0}} \mu^2(d)\tau_3(d) \left| \#\{c \in \mathcal{C} : d \mid (c-r_1)(c-r_2)\cdots(c-r_k)\} - \#\mathcal{C}\prod_{p|d}\frac{\omega_{\mathcal{B}}(p)}{p} \right|,$$

where $\mu(d)$ denotes the Möbius function, $\tau_3(d) := \#\{(u,v,w) \in \mathbb{N}^3 : uvw = d\}$, and

$$L := \sum_{\substack{q \leq Q, \\ p|q \Rightarrow p \in \mathcal{P}_0}} \mu^2(q)\prod_{p|q}\frac{\omega_{\mathcal{B}}(p)}{p - \omega_{\mathcal{B}}(p)}.$$

---

[1]One can also consider the slightly different "finitary" version of the decomposition problem, where we wish to know whether or not $\mathcal{P} \cap [x_0, x] = \mathcal{A} + \mathcal{B}$ for sufficiently large (fixed) x_0 and x. Previously one only had the weaker bound $\max(\#\mathcal{A}, \#\mathcal{B}) \ll x^{1/2}(\log x)^4$ (see [5]) in the finitary version, due to a problem with the ranges of primes for sieving, but our new bound $\ll x^{1/2} \log\log x$ applies in this case too.

Lemma 3.1 follows e.g. from Theorem 7.1 of Friedlander and Iwaniec's book [14] if one chooses $a_n = \sum_{c \in \mathcal{C}} \mathbf{1}_{(c-r_1)(c-r_2)\cdots(c-r_k)=n}$, $P = \prod_{p \in \mathcal{P}_0} p$, and $g(p) = \omega_{\mathcal{B}}(p)/p$ in their theorem.

Lemma 3.2 (Montgomery [25]). *Let $\mathcal{P}$ denote the set of all primes. Let $\mathcal{B}$ denote a set of integers that avoids $\omega_{\mathcal{B}}(p)$ residue classes modulo the prime p. Here $\omega_{\mathcal{B}} : \mathcal{P} \to \mathbb{N}$ with $0 \leq \omega_{\mathcal{B}}(p) \leq p - 1$. Let $\mathcal{B}(x)$ denote the counting function $\mathcal{B}(x) = \sum_{b \leq x, b \in \mathcal{B}} 1$. Then the following upper bound holds, for any Q:*

$$\mathcal{B}(x) \leq \frac{x + Q^2}{L}, \quad \text{where } L = \sum_{q \leq Q} \mu^2(q) \prod_{p | q} \frac{\omega_{\mathcal{B}}(p)}{p - \omega_{\mathcal{B}}(p)}.$$

Lemma 3.3 (Gallagher's larger sieve, [15]). *Let $\mathcal{P}_0$ denote any set of primes such that $\mathcal{A}$ lies modulo p (for $p \in \mathcal{P}_0$) in at most $\nu_{\mathcal{A}}(p)$ residue classes. Then the following bound holds for $\mathcal{A}(N) := \sum_{a \leq N, a \in \mathcal{A}} 1$, provided the denominator is positive:*

$$\mathcal{A}(N) \leq \frac{-\log N + \sum_{p \in \mathcal{P}_0} \log p}{-\log N + \sum_{p \in \mathcal{P}_0} \dfrac{\log p}{\nu_{\mathcal{A}}(p)}}.$$

3.2 The inverse sieve argument

We are going to investigate sums $a + b$ that must consist of restricted prime factors. Suppose that in a particular problem we must have $a + b \not\equiv 0 \bmod p$ whenever $a \in \mathcal{A}$ and $b \in \mathcal{B}$. Then a residue class a that occurs in the set $\mathcal{A}$ forbids a residue class $-a$ in $\mathcal{B}$, in other words (in sieve theory notation, see Lemmas 3.2, 3.3) $\nu_{\mathcal{A}}(p) = \omega_{\mathcal{B}}(p)$. This means that each residue class modulo p can be used for a sieve process: either it occurs in $\mathcal{A}$ and hence $-a \bmod p$ is forbidden in $\mathcal{B}$, which is good when estimating $\mathcal{B}(x)$; or it does not occur in $\mathcal{A}$ and can be used when connecting $\nu_A(p)$ and $\mathcal{A}(x)$. It has proved successful to conclude from $\mathcal{A}(x)$ to $\nu_{\mathcal{A}}(p)$ in some average sense in the "inverse sieve" direction[2], using Gallagher's larger sieve, then to convert this into information on $\omega_{\mathcal{B}}(p)$ on average and use this to give an upper bound on $\mathcal{B}(x)$. (See for example [5, 6, 8, 9]).

It should be emphasized that this kind of argument is quite stable. It works for a wide range of $\mathcal{A}(x)$ and it also works if only a subset of primes can be used for sieving, assuming these primes are somewhat regularly distributed so that the various sums of primes that arise in the argument get a fair share when restricted to the subset.

As a concrete example of (the first part of) this strategy we shall prove the following result, that will be needed several times:

---

[2]The expression "inverse sieve" was introduced in the first author's Habilitationsschrift [7] to recall that in classical sieve problems one deduces bounds on the counting function of a sequence from the number of sifted classes, whereas here one draws conclusions from the size of the counting function to the number of sifted classes.

Lemma 3.4. *Let $\mathcal{P}_0$ be a subset of the primes. Let $\mathcal{A} \subset [1,x]$ be a set of integers that satisfies $\#\mathcal{A} = k \geq 2$, and that occupies $\nu_{\mathcal{A}}(p)$ residue classes modulo the prime p. Then for any $y \geq 10$ we have*

$$\sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \frac{\nu_{\mathcal{A}}(p)}{p} \geq k \sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \frac{1}{p} - \frac{(k^2 - k)\log x}{(y/2)\log(y/2)}. \qquad (1)$$

In particular, if $\sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \log p \geq 8k \log x$ then we have

$$\sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \frac{\nu_{\mathcal{A}}(p)}{p} \geq \frac{k}{2} \sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \frac{1}{p}.$$

Proof of Lemma 3.4. Before we begin the proof, note that we may assume that

$$\sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \log p > \log x,$$

since otherwise we have

$$k \sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \frac{1}{p} \leq \frac{k}{(y/2)\log(y/2)} \sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \log p \leq \frac{k \log x}{(y/2)\log(y/2)}$$

and the assertion (1) of the lemma is trivial.

Applying Gallagher's larger sieve (that is Lemma 3.3), we see

$$-\log x + \sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \frac{\log p}{\nu_{\mathcal{A}}(p)} \leq \frac{-\log x + \sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \log p}{\#\mathcal{A}} = \frac{-\log x + \sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \log p}{k}.$$

Strictly speaking this bound only follows from Lemma 3.3 if the left hand side is strictly positive, but it is trivial otherwise because, by our initial remarks, we may assume that the right hand side is positive. Thus we see

$$\sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \frac{\log p}{\nu_{\mathcal{A}}(p)} - \sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \frac{\log p}{k} \leq (1 - 1/k)\log x.$$

On the other hand, we certainly always have $\nu_{\mathcal{A}}(p) \leq \#\mathcal{A} = k$, so

$$\sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \frac{\log p}{\nu_{\mathcal{A}}(p)} - \sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \frac{\log p}{k} = \sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \frac{(k - \nu_{\mathcal{A}}(p))\log p}{k\nu_{\mathcal{A}}(p)}$$

$$\geq \frac{(y/2)\log(y/2)}{k^2} \sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \frac{(k - \nu_{\mathcal{A}}(p))}{p}.$$

It follows from the inequalities above that

$$\sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \frac{\nu_{\mathcal{A}}(p)}{p} = \sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \frac{k}{p} - \sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \frac{(k - \nu_{\mathcal{A}}(p))}{p}$$

$$\geq \sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \frac{k}{p} - \frac{(k^2 - k)\log x}{(y/2)\log(y/2)}, \qquad (*)$$

9

as claimed.

For the final assertion of the lemma, we just note that if $\sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \log p \geq 8k \log x$ and $y \geq 10$ then

$$k \sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \frac{1}{p} \geq \frac{k}{y \log y} \sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \log p \geq \frac{4k^2 \log x}{y \log(y/2)}.$$

Thus the first term in $(*)$ is at least twice as large as the second. $\square$

We remark that most of the effort in the proof of Lemma 3.4 goes into proving the first assertion, with the error term $2(k^2 - k)(\log x)/(y \log(y/2))$. We will need this for the proofs of Propositions 4.2 and 4.3, below.

3.3 Ruzsa's inequality

Sárközy [34] recently showed that the following inequality of Ruzsa can be successfully applied to questions of ternary decomposability.

Lemma 3.5 (Ruzsa [30], Theorem 9.1). *Let $\mathcal{A}, \mathcal{B}, \mathcal{C}$ be finite sets in a commutative group. We have*

$$|\mathcal{A} + \mathcal{B} + \mathcal{C}|^2 \leq |\mathcal{A} + \mathcal{B}||\mathcal{A} + \mathcal{C}||\mathcal{B} + \mathcal{C}|.$$

Sárközy used this lemma in $\mathbb{F}_p$, but we will see that Ruzsa's inequality provides a simple way to deduce a ternary indecomposability result from a binary counting result over the integers too. For a previous approach to such deductions, which used an inverse sieve and made use of local knowledge modulo p, by means of the Cauchy–Davenport theorem, see [5].

3.4 Finite and infinite decomposability statements

The results and conjectures stated in the introduction and section 2 refer to asymptotic decomposability of infinite sets, but in our proofs we will want to work with finite sets. The reduction to this situation is not difficult, but to avoid any possible obscurity we make some remarks about it now.

Suppose we are investigating whether or not $\mathcal{S} \sim \mathcal{A} + \mathcal{B}$. By definition, this would mean that

$$(\mathcal{A} + \mathcal{B}) \cap [x_0, \infty) = \mathcal{S} \cap [x_0, \infty),$$

where x_0 would be a fixed number depending on $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{S}$. In particular, for any large x we would have

$$\mathcal{S} \cap [x_0, x] \subseteq (\mathcal{A} \cap [0, x]) + (\mathcal{B} \cap [0, x]) \subseteq (\mathcal{S} \cap [x_0, 2x]) \cup [0, x_0].$$

In the problems that we study we know that elements of $\mathcal{S}$ are not divisible by certain primes, and this information can usually be applied straightforwardly to the "finitised" versions. For example, if we are investigating

the smooth numbers $\mathcal{S}_{f(n)}$, where $f(n)$ is an increasing function, then all prime factors larger than $\max(x_0, f(2x))$ are forbidden to occur in the sumset $(\mathcal{A} \cap [0, x]) + (\mathcal{B} \cap [0, x])$, and so can be used for sieving arguments. The other important point is that, for the infinite sets $\mathcal{S}$ that we are interested in, we have

$$\#(\mathcal{S} \cap [x_0, 2x]) \ll \#\mathcal{S} \cap [x_0, x]$$

for all large x, with an absolute implied constant[3]. Thus we know that $(\mathcal{A} \cap [0, x]) + (\mathcal{B} \cap [0, x])$ must make up a positive fraction of $(\mathcal{S} \cap [x_0, 2x]) \cup [0, x_0]$. It is really this assumption, rather than the assumption that they are identical, which is sufficient to make our finitary arguments run, and our general finitary Theorem 4.1 explicitly covers the case where $\mathcal{A} + \mathcal{B}$ is equal to any reasonably large subset of the target set.

If $\mathcal{S} = \mathcal{P}$, the set of primes, then the sieving situation is slightly more complicated because each prime *will* divide precisely one of the elements of $\mathcal{P}$, namely itself. However, if we choose $x \geq x_0^2$, and suppose without loss of generality (for given x) that we have $\#(\mathcal{B} \cap [0, x]) \geq \#(\mathcal{A} \cap [0, x])$, then

$$(\mathcal{A} \cap [0, x]) + (\mathcal{B} \cap [\sqrt{x}, x]) \subseteq \mathcal{P} \cap [\sqrt{x}, 2x].$$

Then all prime factors $p < \sqrt{x}$ are forbidden to occur in the sumset, and our sieving arguments could give an upper bound on $\#(\mathcal{B} \cap [\sqrt{x}, x])$. The reader may check that all the upper bounds we obtain (in section 7) are $\gg \sqrt{x}$, so we have the same bounds (up to a multiplicative constant) for $\#(\mathcal{B} \cap [0, x])$.

In summary, in the following sections we will work with finite sets $\mathcal{S}$, and will investigate the possibility of additive decompositions $\mathcal{S} = \mathcal{A} + \mathcal{B}$, rather than inclusions $\mathcal{A} + \mathcal{B} \subseteq \mathcal{S}$ or asymptotic decompositions $\mathcal{A} + \mathcal{B} \sim \mathcal{S}$. We will also ignore any issues of omitting elements smaller than x_0 or $\sqrt{x}$. The reader may convince themselves that, in view of the above discussion, our results and arguments can easily be tweaked to yield the theorems actually claimed in section 2.

4   A general machinery

4.1   A general additive irreducibility theorem

In this section we will develop some general machinery for obtaining additive irreducibility results, from which the specific theorems in section 2 will follow on inputting suitable information about the specific sets $\mathcal{S}$, such as the smooth numbers, under study, and (to deduce the ternary indecomposability results) on using Ruzsa's inequality.

More precisely, we shall establish the following main result:

---

[3]This is clear when $\mathcal{S}$ is the set of primes, for example. For sets $\mathcal{S}_{f(n)}$ of smooth numbers it is true provided $f(2n) \leq f(n)(1 + (100 \log f(n))/\log n)$, say, as hypothesised in Theorem 2.1 and discussed further in section 5.

Theorem 4.1. *Let x be large, let $\mathcal{P}_0$ be a subset of the primes, and suppose there exists $x^{-1/10} < c \leq 1$ such that*

$$\sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \log p \geq c \sum_{y/2 < p \leq y} \log p, \quad \text{whenever } x^{1/10} \leq y \leq x^{1/2}.$$

Suppose $\mathcal{S} \subseteq [1, x]$ is a non-empty "target set" of integers none of which is divisible by any $p \in \mathcal{P}_0$, and let $\sigma := (\#\mathcal{S})/x$ denote the density of $\mathcal{S}$. Also suppose $\mathcal{S}_0 \subseteq \mathcal{S}$ is any non-empty set, and let $\sigma_0 := (\#\mathcal{S}_0)/(\#\mathcal{S})$ denote the relative density of $\mathcal{S}_0$.

Finally set $K := 1000(\sigma_0 \sigma c^2)^{-1} \log^2 x$ and $\mathcal{P}_0^* := \mathcal{P}_0 \cap [K^3, \infty)$, and suppose that at least one *of the following holds:*

1. *(Sieve Controls Size) we have the lower bound*

$$\sum_{\substack{1 < q \leq \sqrt{x}, \\ p|q \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(q) \prod_{p|q} \frac{2}{p} \geq 10(\sigma_0 \sigma)^{-1};$$

2. *(Bombieri–Vinogradov) there exists a positive integer Q such that*

$$\sum_{\substack{1 < q \leq Q, \\ p|q \Rightarrow p \in \mathcal{P}_0^*}} \frac{\mu^2(q)}{q} \geq 10\sigma_0^{-1},$$

and also

$$\sum_{\substack{d \leq Q^2, \\ p|d \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(d) \tau_3(d)^{1+(\log K)/(\log 3)} \max_{(a,d)=1} \left| \#\{s \in \mathcal{S} : s \equiv a \bmod d\} - \frac{\#\mathcal{S}}{\phi(d)} \right| \leq \frac{\#\mathcal{S}}{2K\sigma_0^{-1}},$$

where $\mu(d)$ denotes the Möbius function, $\phi(d)$ denotes the Euler totient function, and $\tau_3(d) := \#\{(u, v, w) \in \mathbb{N}^3 : uvw = d\}$.

Then if there is a decomposition $\mathcal{S}_0 = \mathcal{A} + \mathcal{B}$, for some sets $\mathcal{A}, \mathcal{B}$ of non-negative integers with $2 \leq \#\mathcal{A} \leq \#\mathcal{B}$, one has the upper bound $\#\mathcal{B} \ll (\sqrt{x} \log^4 x)/c^4$, and therefore also the lower bound $\#\mathcal{A} \gg (\sqrt{x} \sigma_0 \sigma c^4)/\log^4 x$.

We remark that the reader will not lose much by thinking of Theorem 4.1 in the simpler case where $\mathcal{S}_0 = \mathcal{S}$, and therefore $\sigma_0 = 1$. (And for simplicity we will assume this in all of our applications.) We allow ourselves the possibility of handling a subset $\mathcal{S}_0$ because, as discussed in section 3.4, that is the situation that arises when carefully turning asymptotic irreducibility questions into finitary questions.

Note that the lower bound demanded in the "Sieve Controls Size" condition involves a factor $(\sigma_0 \sigma)^{-1}$, whereas the lower bound demanded in the

"Bombieri–Vinogradov condition" only involves the smaller quantity σ_0^{-1} (but in the Bombieri–Vinogradov condition one also needs an upper bound for another sum, as in part 2 of the theorem above). The "Sieve Controls Size" condition implies that one can more-or-less extract $\mathcal{S}_0$ by sieving the integers less than x by the primes $p \in \mathcal{P}_0^*$, and ultimately that one can run a sieve process *on all those integers* to bound $\#\mathcal{A}$ and $\#\mathcal{B}$. If the "Bombieri–Vinogradov condition" is satisfied then one may not be able to succeed using a sieve process on $[1, x]$, but instead one can run a sieve process *inside* $\mathcal{S}$, using Selberg's sieve and other ingredients, to bound $\#\mathcal{A}$ and $\#\mathcal{B}$.

As we will see in sections 5 – 6, Theorem 4.1 can be applied to sets of smooth numbers because they satisfy the "Bombieri–Vinogradov condition", and it can be applied to semigroup-type sets $\mathcal{S}$, whose elements are generated multiplicatively by a dense subsequence of the primes, because they satisfy the "Sieve Controls Size" condition. One could also use Theorem 4.1 to study the primes themselves, which satisfy both conditions.

We will explain and prove Theorem 4.1 in the next subsection, on the assumption of some sieve theory type propositions, and then in section 4.3 we will prove those propositions. We hope that establishing a general result will clarify, as well as extend, the approach of previous papers[4] such as [5].

4.2   Outline proof of Theorem 4.1

Suppose there is an additive decomposition $\mathcal{S}_0 = \mathcal{A} + \mathcal{B}$, where $2 \le \#\mathcal{A} \le \#\mathcal{B}$. If we let $k := \#\mathcal{A}$, then since $\#(\mathcal{A} + \mathcal{B}) \le k \cdot \#\mathcal{B}$ we must certainly have

$$\#\mathcal{B} \ge \frac{\#\mathcal{S}_0}{k} = \frac{\sigma_0 \sigma x}{k}.$$

On the other hand, if the elements of $\mathcal{S}_0 \subseteq \mathcal{S}$ are not divisible by certain primes then, for given $\mathcal{A}$, there will be congruence constraints modulo those primes on the possible elements of $\mathcal{B}$. We will use various sieve arguments to place upper bounds on $\#\mathcal{B}$, which will contradict the lower bound $\#\mathcal{B} \ge (\#\mathcal{S}_0)/k$ unless k and $\#\mathcal{B}$ are both "fairly close" to $\sqrt{x}$ (as asserted in Theorem 4.1).

In all of the results stated in this subsection, $\mathcal{S} \subseteq [1, x]$, $\mathcal{P}_0$ and $\mathcal{P}_0^* = \mathcal{P}_0 \cap [K^3, \infty)$ will denote subsets of the integers and the primes, as in the statement of Theorem 4.1, and $0 \le a_1 < a_2 < \ldots < a_k \le x$ will denote any distinct integers.

Firstly we present two results, corresponding to the "Sieve Controls Size" condition and the "Bombieri–Vinogradov condition", that can be useful when k is assumed to be fairly small:

---

[4]The previous work all roughly corresponds to various special applications of the "Sieve Controls Size" part of Theorem 4.1.

Proposition 4.2. *For any $2 \leq k \leq K = 1000(\sigma_0 \sigma c^2)^{-1} \log^2 x$ we have the upper bound*

$$\#\{s \in \mathcal{S} : s \not\equiv a_i \mod p, \quad 1 \leq i \leq k, \, p \in \mathcal{P}_0^*\} \leq \frac{4x}{(k/2) \sum_{\substack{1 < q \leq \sqrt{x}, \\ p|q \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(q) \prod_{p|q} \frac{2}{p}}.$$

Proposition 4.3. *For any $2 \leq k \leq K = 1000(\sigma_0 \sigma c^2)^{-1} \log^2 x$, and any positive integer parameter Q, $\#\{s \in \mathcal{S} : s \not\equiv a_i \mod p, \quad 1 \leq i \leq k, \, p \in \mathcal{P}_0^*\}$ is at most*

$$\frac{2\#\mathcal{S}}{(k-1) \sum_{\substack{1 < q \leq Q, \\ p|q \Rightarrow p \in \mathcal{P}_0^*}} \frac{\mu^2(q)}{q}} + \sum_{\substack{d \leq Q^2, \\ p|d \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(d) \tau_3(d)^{1 + \frac{\log k}{\log 3}} \max_{(a,d)=1} \left| \#\{s \in \mathcal{S} : s \equiv a \mod d\} - \frac{\#\mathcal{S}}{\phi(d)} \right|.$$

At first glance it might seem that the case when k is small should not be difficult to handle. However, the situation is a little delicate, since there (probably) *are* sets $\mathcal{S}$, very like the ones that we shall study, that admit decompositions as $\mathcal{A} + \mathcal{B}$ with $\#\mathcal{A} \geq 2$ small. For example, it is conjectured that the twin primes constitute roughly a proportion $1/\log x$ of all the primes, and they clearly admit a decomposition with $\mathcal{A} = \{0, 2\}$. Also, the quadruples of primes $\{10k + 1, 3, 7, 9\}$ even admit a ternary decomposition $\{1, 3\} + \{0, 6\} + \mathcal{C}$. Here $\mathcal{C}$ is expected to be infinite. Thus some work is really required, as in the above propositions, to know that we can perform useful sieving moves to handle small k and get our argument started.

The next result will provide our main estimates when k is larger: however, we state it only as a lemma in the first instance, since it will be incorporated into a slightly more general proposition immediately below.

Lemma 4.4. *Let k be any natural number, and suppose that*

$$y_1 < \frac{y_2}{2} < y_2 < \frac{\sqrt{x}}{y_1}$$

are large and are such that

$$\sum_{y_1/2 < p \leq y_1, p \in \mathcal{P}_0} \log p \geq 8k \log x \quad and \quad \sum_{y_2/2 < p \leq y_2, p \in \mathcal{P}_0} \log p \geq 8k \log x.$$

Then $\#\{s \in \mathcal{S} : s \not\equiv a_i \mod p, \quad 1 \leq i \leq k, \, p \in \mathcal{P}_0\}$ is

$$\leq \frac{128x \log y_1 \log y_2}{k^2} \cdot \frac{\sum_{y_1/2 < p \leq y_1} \log p}{\sum_{y_1/2 < p \leq y_1, p \in \mathcal{P}_0} \log p} \cdot \frac{\sum_{y_2/2 < p \leq y_2} \log p}{\sum_{y_2/2 < p \leq y_2, p \in \mathcal{P}_0} \log p}.$$

The slight problem with Lemma 4.4 is that if k is very large, e.g. larger than about $x^{1/4}$, then it is not possible to find values y_1, y_2 satisfying the hypotheses. However, since the bound in the lemma decays quadratically with the number of values a_i we can still obtain strong information if we artificially reduce the value of k, i.e. only work with a subset of the a_i. The following result, which also includes Lemma 4.4, records the kind of bound obtained:

Proposition 4.5. *Suppose that $y_1 < y_2/2 < y_2 < \sqrt{x}/y_1$ are large and are such that*

$$\sum_{y_1/2 < p \leq y_1, p \in \mathcal{P}_0} \log p \geq 8 \log x \quad and \quad \sum_{y_2/2 < p \leq y_2, p \in \mathcal{P}_0} \log p \geq 8 \log x,$$

and let k be arbitrary. Then $\#\{s \in \mathcal{S} : s \not\equiv a_i \mod p, \quad 1 \leq i \leq k, \ p \in \mathcal{P}_0\}$ is

$$\leq 128 x M \log y_1 \log y_2 \cdot \frac{\sum_{y_1/2 < p \leq y_1} \log p}{\sum_{y_1/2 < p \leq y_1, p \in \mathcal{P}_0} \log p} \cdot \frac{\sum_{y_2/2 < p \leq y_2} \log p}{\sum_{y_2/2 < p \leq y_2, p \in \mathcal{P}_0} \log p},$$

where

$$M = M(y_1, y_2, x, k, \mathcal{P}_0) := \max\left\{\frac{1}{k^2}, \frac{256 \log^2 x}{(\sum_{y_1/2 < p \leq y_1, p \in \mathcal{P}_0} \log p)^2}, \frac{256 \log^2 x}{(\sum_{y_2/2 < p \leq y_2, p \in \mathcal{P}_0} \log p)^2}\right\}.$$

The foregoing propositions can be applied to prove Theorem 4.1, as follows:

- by subtracting a suitable integer from all the elements of $\mathcal{A}$, and adding it to all the elements of $\mathcal{B}$, we may assume that we have $\mathcal{S}_0 = \mathcal{A} + \mathcal{B}$ for some set $\mathcal{A}$ of non-negative integers that contains zero. Note this means, in particular, that $\mathcal{B} = \{0\} + \mathcal{B}$ is a subset of $\mathcal{S}_0$, and therefore a subset of $\mathcal{S}$, and in fact

$$\mathcal{B} \subseteq \{s \in \mathcal{S} : s \not\equiv -a \mod p, \quad a \in \mathcal{A}, \ p \in \mathcal{P}_0\}.$$

- recall that we write $k = \#\mathcal{A}$.

- if $2 \leq k \leq 1000(\sigma_0 \sigma c^2)^{-1} \log^2 x$, and the "Sieve Controls Size" condition is satisfied, then the bound in Proposition 4.2 is $\leq 4x/(5k\sigma_0^{-1}\sigma^{-1})$, which is $< x\sigma_0\sigma/k = (\#\mathcal{S}_0)/k$. This contradicts our earlier observation that $\#\mathcal{B} \geq (\#\mathcal{S}_0)/k$, so we conclude that k cannot be so small.

- if $2 \leq k \leq 1000(\sigma_0 \sigma c^2)^{-1} \log^2 x$, and the "Bombieri–Vinogradov condition" is satisfied, then the bound in Proposition 4.3 is $\leq (2\#\mathcal{S})/(10\sigma_0^{-1}(k-1)) + (\#\mathcal{S})/(2K\sigma_0^{-1})$, which is at most $(9\#\mathcal{S})/(10\sigma_0^{-1}k) = (9\#\mathcal{S}_0)/(10k)$. Again, $\mathcal{B}$ cannot be this small and so k cannot be so small.

- if $k > 1000(\sigma_0 \sigma c^2)^{-1} \log^2 x$, we apply Proposition 4.5 with the choices $y_1 = x^{1/4}/3$ and $y_2 = x^{1/4}$, say. We have $\sum_{y_1/2 < p \le y_1, p \in \mathcal{P}_0} \log p \ge c \sum_{y_1/2 < p \le y_1} \log p \ge cx^{1/4}/10$, say (since x, and therefore y_1 and y_2, are large), similarly for the sum on the range $y_2/2 < p \le y_2$, so that

$$\begin{aligned} \#\mathcal{B} &\le \#\{s \in \mathcal{S} : s \not\equiv -a \mod p, \quad a \in \mathcal{A}, \ p \in \mathcal{P}_0\} \\ &\le 128x \max\{\frac{1}{k^2}, \frac{25600 \log^2 x}{c^2 \sqrt{x}}\} \frac{\log^2 x}{c^2}. \end{aligned}$$

Now we cannot have $\#\mathcal{B} \le (128x \log^2 x)/(k^2 c^2)$, since this bound is $< (x\sigma_0\sigma)/k$ when k is $> 1000(\sigma_0\sigma c^2)^{-1} \log^2 x$, and we must have $\#\mathcal{B} \ge (\#\mathcal{S}_0)/k = x\sigma_0\sigma/k$. Therefore we must have

$$\#\mathcal{B} \ll x \cdot \frac{\log^2 x}{c^2 \sqrt{x}} \cdot \frac{\log^2 x}{c^2} = \frac{\sqrt{x} \log^4 x}{c^4},$$

as asserted in Theorem 4.1. Since we have $\#\mathcal{A}\#\mathcal{B} \ge \#\mathcal{S}_0 = x\sigma_0\sigma$, the lower bound claimed for $\#\mathcal{A}$ also immediately follows.

$\square$

4.3   Proofs of the propositions

Proof of Proposition 4.2. For each $p \in \mathcal{P}_0^*$ let $1 \le \nu(p) \le k$ denote the number of distinct residue classes modulo p occupied by the numbers $a_1, a_2, \dots, a_k$. Then in view of Montgomery's Large Sieve (that is Lemma 3.2), applied with the choice $Q = \sqrt{x}$, to prove Proposition 4.2 it will suffice to show that

$$\sum_{\substack{q \le \sqrt{x}, \\ p|q \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(q) \prod_{p|q} \frac{\nu(p)}{p - \nu(p)} \ge (1/2) \cdot (k/2) \sum_{\substack{1 < q \le \sqrt{x}, \\ p|q \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(q) \prod_{p|q} \frac{2}{p},$$

provided $2 \le k \le K$. Recall that if $p \in \mathcal{P}_0^*$ then $p \ge K^3$, by definition.

Now note that, for any q counted in the sum on the left hand side,

$$\prod_{p|q} \frac{\nu(p)}{p - \nu(p)} = \prod_{p|q} \frac{(\nu(p)/p)}{1 - (\nu(p)/p)} = \left(\prod_{p|q} \frac{\nu(p)}{p} \right) \left(\prod_{p|q} \sum_{i=0}^{\infty} \left(\frac{\nu(p)}{p} \right)^i \right),$$

so if we define a completely multiplicative function $f(n)$ by setting $f(p) = \nu(p) \mathbf{1}_{p \in \mathcal{P}_0^*}$, then

$$\sum_{\substack{q \le \sqrt{x}, \\ p|q \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(q) \prod_{p|q} \frac{\nu(p)}{p - \nu(p)} = \sum_{q \le \sqrt{x}} \mu^2(q) \frac{f(q)}{q} \sum_{\substack{r=1, \\ p|r \Rightarrow p|q}}^{\infty} \frac{f(r)}{r} \ge \sum_{n \le \sqrt{x}} \frac{f(n)}{n}.$$

Let us also define a completely multiplicative function $g(n)$, by setting $g(p) = k\mathbf{1}_{p \in \mathcal{P}_0^*}$. Thus $0 \leq f(p) \leq g(p) < p$ for each prime p. It is a general fact, whose proof we shall reproduce in just a moment, that for any non-negative completely multiplicative functions related in this way,

$$\sum_{n \leq \sqrt{x}} \frac{f(n)}{n} \geq \prod_{p \leq \sqrt{x}} \left(1 - \frac{g(p)}{p}\right) \prod_{p \leq \sqrt{x}} \left(1 - \frac{f(p)}{p}\right)^{-1} \sum_{n \leq \sqrt{x}} \frac{g(n)}{n}. \qquad (**)$$

Thus we have

$$\sum_{\substack{q \leq \sqrt{x}, \\ p|q \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(q) \prod_{p|q} \frac{\nu(p)}{p - \nu(p)} \geq \sum_{n \leq \sqrt{x}} \frac{f(n)}{n} \geq \prod_{\substack{p \leq \sqrt{x}, \\ p \in \mathcal{P}_0^*}} \left(1 - \frac{(k - \nu(p))}{p - \nu(p)}\right) \sum_{\substack{q \leq \sqrt{x}, \\ p|q \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(q) \prod_{p|q} \frac{k}{p},$$

noting that the final sum over q just consists of the squarefree terms from the corresponding sum $\sum_{n \leq \sqrt{x}} \frac{g(n)}{n}$.

Now if $p \in \mathcal{P}_0^*$ we certainly have $p \geq K^3 \geq 2k \geq 2\nu(p)$, and therefore

$$0 \leq \frac{k - \nu(p)}{p - \nu(p)} \leq \frac{2(k - \nu(p))}{p}.$$

Then if we use the fact that $1 - t \geq e^{-2t}$ when $0 \leq t \leq 1/4$, and apply Lemma 3.4 on dyadic intervals (starting at K^3), we see the product over primes in the previous display is

$$\prod_{\substack{p \leq \sqrt{x}, \\ p \in \mathcal{P}_0^*}} \left(1 - \frac{(k - \nu(p))}{p - \nu(p)}\right) \geq e^{-4 \sum_{p \leq \sqrt{x}, p \in \mathcal{P}_0^*} \frac{k - \nu(p)}{p}}$$

$$\geq e^{-(8(k^2 - k) \log x)/(K^3 \log(K^3))} \geq e^{-(8 \log x)/K} \geq 1/2,$$

say. Finally we note that

$$\sum_{\substack{q \leq \sqrt{x}, \\ p|q \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(q) \prod_{p|q} \frac{k}{p} > \sum_{\substack{1 < q \leq \sqrt{x}, \\ p|q \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(q) \prod_{p|q} \frac{k}{p} \geq (k/2) \sum_{\substack{1 < q \leq \sqrt{x}, \\ p|q \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(q) \prod_{p|q} \frac{2}{p},$$

since each term in the second sum involves at least one factor $k = (k/2) \cdot 2$. This all suffices to prove Proposition 4.2.

It remains to prove the general inequality $(**)$, where f and g are completely multiplicative functions satisfying $0 \leq f(p) \leq g(p) < p$. The reader may find this result in the forthcoming book [16] of Granville and Soundararajan, but we will reproduce the short proof here. Let h be the multiplicative function defined by

$$h(p^b) = (g(p) - f(p))g(p^{b-1}), \quad p \leq \sqrt{x}, \quad b \geq 1,$$

17

and by $h(p^b) = 0$ when $p > \sqrt{x}$. Thus $h \geq 0$, and the reader may check that $g = f * h$, where $*$ here denotes Dirichlet convolution. Then

$$\sum_{N \leq \sqrt{x}} \frac{g(N)}{N} = \sum_{mn \leq \sqrt{x}} \frac{h(m)f(n)}{mn} \leq \left(\sum_{m=1}^{\infty} \frac{h(m)}{m} \right) \sum_{n \leq \sqrt{x}} \frac{f(n)}{n}$$

$$= \prod_{p \leq \sqrt{x}} \left(1 - \frac{g(p)}{p} \right)^{-1} \prod_{p \leq \sqrt{x}} \left(1 - \frac{f(p)}{p} \right) \sum_{n \leq \sqrt{x}} \frac{f(n)}{n},$$

as claimed in $(**)$. $\qquad\square$

Proof of Proposition 4.3. The proof is like that of Proposition 4.2, but crucially using Selberg's sieve instead of the large sieve. Let $0 \leq \nu(p) \leq k$ denote the number of *non-zero* residue classes modulo p occupied by the numbers $a_1, a_2, \ldots, a_k$, noting this is a slightly different definition than previously. Applying Lemma 3.1 with the choice $\omega_{\mathcal{B}}(p) = \frac{p\nu(p)}{p-1}$, (and $\mathcal{P}_0$ replaced by $\mathcal{P}_0^*$, and $\mathcal{C}$ replaced by $\mathcal{S}$, and recalling that $p \geq K^3$ if $p \in \mathcal{P}_0^*$), we find $\#\{s \in \mathcal{S} : s \not\equiv a_i \mod p, \quad 1 \leq i \leq k, \ p \in \mathcal{P}_0^*\}$ is at most

$$\frac{\#\mathcal{S}}{\sum_{\substack{q \leq Q, \\ p|q \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(q) \prod_{p|q} \frac{\nu(p)}{p - \nu(p)}} + \sum_{\substack{d \leq Q^2, \\ p|d \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(d)\tau_3(d) \left| \#\{s \in \mathcal{S} : d \mid \prod_{i=1}^{k}(s - a_i)\} - \#\mathcal{S} \prod_{p|d} \frac{\nu(p)}{(p-1)} \right|.$$

Arguing exactly as in the proof of Proposition 4.2, except choosing $g(p) = k - 1$ (instead of k) for those primes $p \in \mathcal{P}_0^*$ such that one of $a_1, \ldots, a_k$ is congruent to zero modulo p, one finds that

$$\sum_{\substack{q \leq Q, \\ p|q \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(q) \prod_{p|q} \frac{\nu(p)}{p - \nu(p)} \geq \frac{(k-1)}{2} \sum_{\substack{1 < q \leq Q, \\ p|q \Rightarrow p \in \mathcal{P}_0^*}} \frac{\mu^2(q)}{q}.$$

Thus the first term in the previous display is at most as large as the first term in the statement of Proposition 4.3. It remains to treat the second term.

But a squarefree integer d divides $\prod_{i=1}^{k}(s - a_i)$ if and only if, for each prime divisor p of d, s lies modulo p in one of the residue classes occupied by the numbers $a_1, \ldots, a_k$. By assumption, $s \not\equiv 0 \mod p$ for any $s \in \mathcal{S}$ and $p \in \mathcal{P}_0^*$, so actually s must lie in one of the $\nu(p)$ non-zero residue classes occupied by $a_1, \ldots, a_k$. By the Chinese Remainder Theorem, this is the same as saying that s lies modulo d in one of $\prod_{p|d} \nu(p)$ residue classes, so the triangle inequality implies the second term is

$$\leq \sum_{\substack{d \leq Q^2, \\ p|d \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(d)\tau_3(d) \left(\prod_{p|d} \nu(p) \right) \max_{(a,d)=1} \left| \#\{s \in \mathcal{S} : s \equiv a \mod d\} - \frac{\#\mathcal{S}}{\phi(d)} \right|.$$

18

Since $\prod_{p|d} \nu(p) \leq \prod_{p|d} k = (\prod_{p|d} 3)^{(\log k)/(\log 3)}$, and $\prod_{p|d} 3$ is at most $\tau_3(d)$, the bound asserted in Proposition 4.3 follows. $\square$

Proof of Lemma 4.4. This is a consequence of Montgomery's Large Sieve and an "inverse sieve" argument, similar to (but easier than) Proposition 4.2. For each $p \in \mathcal{P}_0$, let $1 \leq \nu(p) \leq k$ denote the number of distinct residue classes modulo p occupied by the numbers $a_1, a_2, \ldots, a_k$. Using Lemma 3.2 with the choice $Q = \sqrt{x}$, we see $\#\{s \in \mathcal{S} : s \not\equiv a_i \mod p, \quad 1 \leq i \leq k, \ p \in \mathcal{P}_0\}$ is

$$\leq \frac{2x}{\sum_{\substack{q \leq \sqrt{x}, \\ p|q \Rightarrow p \in \mathcal{P}_0}} \mu^2(q) \prod_{p|q} \frac{\nu(p)}{p - \nu(p)}} \leq \frac{2x}{(\sum_{y_1/2 < p \leq y_1, p \in \mathcal{P}_0} \frac{\nu(p)}{p - \nu(p)})(\sum_{y_2/2 < p \leq y_2, p \in \mathcal{P}_0} \frac{\nu(p)}{p - \nu(p)})},$$

since each product $p_1 p_2$ with $y_1/2 < p_1 \leq y_1$ and $y_2/2 < p_2 \leq y_2$ is a squarefree integer that is $\leq \sqrt{x}$.

The above is at most $2x/((\sum_{y_1/2 < p \leq y_1, p \in \mathcal{P}_0} \nu(p)/p)(\sum_{y_2/2 < p \leq y_2, p \in \mathcal{P}_0} \nu(p)/p))$, and by Lemma 3.4 each of the terms in the denominator is at least

$$\frac{k}{2} \sum_{y_i/2 < p \leq y_i, p \in \mathcal{P}_0} \frac{1}{p} \geq \frac{k}{2 y_i \log y_i} \sum_{y_i/2 < p \leq y_i, p \in \mathcal{P}_0} \log p \geq \frac{k}{4 \log y_i} \frac{\sum_{y_i/2 < p \leq y_i, p \in \mathcal{P}_0} \log p}{\sum_{y_i/2 < p \leq y_i} \log p} \sum_{y_i/2 < p \leq y_i} \frac{\log p}{p}.$$

Lemma 4.4 follows on noting that, by standard results on the distribution of prime numbers, $\sum_{y_i/2 < p \leq y_i} (\log p)/p \geq 1/2$ when y_i is large. $\square$

Proof of Proposition 4.5. Define

$$k_0 := \min\{k, \lfloor (\sum_{y_1/2 < p \leq y_1, p \in \mathcal{P}_0} \log p)/8 \log x \rfloor, \lfloor (\sum_{y_2/2 < p \leq y_2, p \in \mathcal{P}_0} \log p)/8 \log x \rfloor\},$$

where $\lfloor \cdot \rfloor$ denotes the integer part, noting that $k_0 \geq 1$ by the hypotheses of Proposition 4.5. Then we trivially have that

$$\#\{s \in \mathcal{S} : s \not\equiv a_i \mod p, \quad 1 \leq i \leq k, \ p \in \mathcal{P}_0\} \leq \#\{s \in \mathcal{S} : s \not\equiv a_i \mod p, \quad 1 \leq i \leq k_0, \ p \in \mathcal{P}_0\},$$

and Proposition 4.5 follows if we use Lemma 4.4 to bound the right hand side (and observe that $\lfloor t \rfloor^2 \geq t^2/4$ for all $t \geq 1$). $\square$

5 Proofs of Theorem 2.1 and Corollary 2.2, application to smooth numbers

We shall prove Theorem 2.1 first, by feeding suitable information about the set of y-smooth numbers into our general additive irreducibility result, Theorem 4.1. Corollary 2.2 will follow from Theorem 2.1 by a general type of argument using Ruzsa's inequality.

More specifically, for any $x, y \geq 2$ and any positive integers a, d let us recall the following standard notation for smooth numbers:

$$\Psi(x, y) := \mathcal{S}_y(x) = \#\{n \leq x : p \nmid n, \; p > y\},$$

$$\Psi_d(x, y) := \#\{n \leq x : (n, d) = 1, \; p \nmid n, \; p > y\},$$

$$\Psi(x, y; a, d) := \#\{n \leq x : n \equiv a \bmod d, \; p \nmid n, \; p > y\}.$$

We shall require the following two results: a lower bound, and a Bombieri–Vinogradov type result.

Theorem 5.1 (See e.g. Corollary 7.9 of Montgomery and Vaughan [26])**.**
For any fixed $a \geq 1$ and any $\log^a x \leq y \leq x$ we have

$$\Psi(x, y) \geq \Psi(x, \log^a x) = x^{1 - 1/a + o_a(1)},$$

where the term $o_a(1)$ tends to 0 as $x \to \infty$.

Theorem 5.2 (See Proposition 2 of Harper [19])**.** *Let $0 < \eta \leq 1/80$ be any fixed constant, and let $\mathcal{P}_0 \subseteq [x^\eta, \infty)$ be a set of primes. Then for any large $y \leq x$ and any $x^\eta \leq Q \leq \sqrt{x}$ we have*

$$\sum_{\substack{d \leq Q, \\ p | d \Rightarrow p \in \mathcal{P}_0}} \max_{(a, d) = 1} \left| \Psi(x, y; a, d) - \frac{\Psi_d(x, y)}{\phi(d)} \right| \ll \log^{7/2} x \sqrt{\Psi(x, y)} (Q + x^{1/2 - \eta} \log^2 x),$$

where $\phi(d)$ is Euler's totient function.

Theorem 5.2 follows on noting that the left hand side is at most

$$\sum_{\substack{d \leq Q, \\ p | d \Rightarrow p \in \mathcal{P}_0}} \frac{1}{\phi(d)} \sum_{\substack{\chi \, (\mathrm{mod}\ d), \\ \chi \neq \chi_0}} \left| \sum_{\substack{n \leq x, \\ n \in \mathcal{S}_y}} \chi(n) \right| \leq \sum_{x^\eta \leq r \leq Q} \sum_{\substack{\chi^* \, (\mathrm{mod}\ r), \\ \chi^* \text{ primitive}}} \sum_{x^\eta \leq d \leq Q} \frac{1}{\phi(d)} \sum_{\substack{\chi \, (\mathrm{mod}\ d), \\ \chi^* \text{ induces } \chi}} \left| \sum_{\substack{n \leq x, \\ n \in \mathcal{S}_y}} \chi(n) \right|,$$

and then using Proposition 2 of Harper [19] (and the discussion in the two paragraphs following Proposition 2) to bound this. Note that because we are only interested in moduli d with all their prime factors larger than x^η, we do not need to include values r of the conductor that are smaller than x^η, which allows us to give quite a strong bound. See also Fouvry and Tenenbaum's paper [13] for an earlier Bombieri–Vinogradov type result for smooth numbers, which would have allowed us to prove an additive irreducibility result on the smaller range $x^{(\log \log \log x) / \log \log x} \leq y \leq x^\kappa$.

Note that the counting function $\Psi(x, y)$ is very small when y is small (and in fact even when y is quite large), but that the set of prime factors forbidden to occur in the y-smooth numbers, namely the primes between y and x, is not a very large set as measured by the kinds of denominators occurring in sieve bounds. In particular, we cannot satisfy the "Sieve Controls Size" condition, or anything similar, in this case, but instead we can use Theorem 5.2 to satisfy the "Bombieri–Vinogradov condition" in Theorem 4.1.

Proof of Theorem 2.1. Recall from the statement that κ is a sufficiently small absolute constant, and D is a sufficiently large absolute constant.

We will apply Theorem 4.1 to the set $\mathcal{S} = \mathcal{S}_y \cap [0, x]$, where we put $y = f(x)$, aiming to deduce[5] that if $\mathcal{S} = \mathcal{A} + \mathcal{B}$ then we must have $\#\mathcal{A}, \#\mathcal{B} \ll \sqrt{x} \log^4 x$.

Since, by assumption, we have $\log^D x \leq y \leq x^\kappa \leq x^{1/10}/2$ (say), we can take $\mathcal{P}_0 = \{x^\kappa < p \leq x : p \text{ prime}\}$ and $c = 1$ in Theorem 4.1. Moreover, using Theorem 5.1 we see

$$\sigma = \frac{\Psi(x, y)}{x} \geq \frac{\Psi(x, \log^D x)}{x} = x^{-1/D + o(1)},$$

which is at least $1000 x^{-2/D} \log^2 x$ (say) provided x is large enough. Thus we have $K = 1000 \sigma^{-1} \log^2 x \leq x^{2/D}$ in Theorem 4.1, so that $\mathcal{P}_0^* = \mathcal{P}_0$ in this case (provided x is large enough and $6/D \leq \kappa$, which we may certainly assume).

Now it suffices to check that the two parts of the "Bombieri–Vinogradov condition" in Theorem 4.1 are satisfied. We choose $Q = x^{1/5}$ and note that, in view of a standard result on the distribution of prime numbers (and since κ is small),

$$\sum_{\substack{1 < q \leq Q, \\ p|q \Rightarrow p \in \mathcal{P}_0^*}} \frac{\mu^2(q)}{q} \geq \sum_{x^\kappa < p \leq x^{1/5}} \frac{1}{p} = \log\log(x^{1/5}) - \log\log(x^\kappa) + O(1) \geq 10,$$

as required.

Next, if $d \leq Q^2 = x^{2/5}$ has all of its prime factors from $\mathcal{P}_0^* = \mathcal{P}_0$ then it has at most $\log(x^{2/5})/\log(x^\kappa) < \kappa^{-1}$ prime factors, and therefore

$$\tau_3(d) \leq 3^{1/\kappa},$$

since $\tau_3(d)$ is at most the number of ways of partitioning the prime factors

---

[5]Recall the discussion of finite and infinite decomposability statements in section 3.4. Since we assume in Theorem 2.1 that $f(n)$ is increasing, $f(2n) \leq f(n)(1 + (100 \log f(n))/\log n)$, and that $f(n) \geq \log^2 n$, it is a fact that

$$\#\mathcal{S}_{f(n)} \cap [0, x] \geq \#\mathcal{S}_{f(x/2)} \cap [x/2, x] \gg \#\mathcal{S}_{f(x/2)} \cap [0, x] \gg \#\mathcal{S}_y \cap [0, x],$$

with an absolute implied constant. Thus we can deduce Theorem 2.1 from results about the finitary case $\mathcal{S} = \mathcal{A} + \mathcal{B}$. (For the final inequality see the proof of Smooth Numbers Result 2 in [19], which can be adapted to show that if $y \geq \log^2 x$, say, then $\Psi(x, y(1 + (100 \log y)/\log x)) \ll \Psi(x, y)$ (since if $y \geq \log^2 x$ then $y/\log x \geq \sqrt{y}$, so one can save a factor of $\log y$ in the estimation of $\sum_{y < p \leq y(1 + (100 \log y)/\log x)}$ there).)

of d into three sets. Thus

$$\sum_{\substack{d \leq Q^2, \\ p|d \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(d) \tau_3(d)^{1+(\log K)/(\log 3)} \max_{(a,d)=1} \left| \Psi(x,y;a,d) - \frac{\Psi(x,y)}{d} \right|$$

$$\leq K^C \sum_{\substack{d \leq x^{2/5}, \\ p|d \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(d) \max_{(a,d)=1} \left| \Psi(x,y;a,d) - \frac{\Psi_d(x,y)}{\phi(d)} \right|$$

for a certain constant $C = C(\kappa)$, using the fact that $\Psi(x,y) = \Psi_d(x,y)$ if all the prime factors of d are larger than y. Recalling that we have $K \leq x^{2/D}$, where D is suitably large in terms of κ, then Theorem 5.2 shows the above is $\leq \Psi(x,y)/2K$, and so the Bombieri–Vinogradov condition is satisfied. $\quad\square$

Proof of Corollary 2.2. Suppose, for a contradiction, that we had $\mathcal{S}_y \cap [0,x] = \mathcal{A} + \mathcal{B} + \mathcal{C}$ for sets $\mathcal{A}, \mathcal{B}, \mathcal{C}$ of positive integers each containing at least two elements. By Ruzsa's inequality, (that is Lemma 3.5),

$$\Psi(x,y)^2 = \mathcal{S}_y(x)^2 \leq (\#(\mathcal{A} + \mathcal{B}))(\#(\mathcal{A} + \mathcal{C}))(\#(\mathcal{B} + \mathcal{C})).$$

But if $\mathcal{S}_y \cap [0,x] = \mathcal{A} + \mathcal{B} + \mathcal{C} = (\mathcal{A} + \mathcal{B}) + \mathcal{C}$ then, by Theorem 2.1, $\#(\mathcal{A} + \mathcal{B}) \ll \sqrt{x} \log^4 x$, and similarly we would have $\#(\mathcal{A} + \mathcal{C}), \#(\mathcal{B} + \mathcal{C}) \ll \sqrt{x} \log^4 x$. Thus

$$\Psi(x,y)^2 \ll x^{3/2} \log^{12} x,$$

which massively contradicts our lower bound $\Psi(x,y) \geq \Psi(x, \log^D x) = x^{1-1/D+o(1)}$ provided D and x are large enough. $\quad\square$

6   Proofs of Theorem 2.4 and Corollary 2.5, sets generated by a dense subsequence of the primes

In this section we shall prove Theorem 2.4 and Corollary 2.5.

We first state Wirsing's mean value theorem, as this is used in counting the elements in $\mathcal{Q}(\mathcal{T})$.

Lemma 6.1 (Wirsing [40])**.** *Let f be a non-negative multiplicative arithmetic function that satisfies the following hypotheses:*

1.
$$\sum_{p \leq x} f(p) \frac{\log p}{p} \sim \tau \log x,$$

   where τ is a positive constant.

2. *$f(p) = O(1)$.*

3.
$$\sum_{p \text{ prime, } n \geq 2} \frac{f(p^n)}{p^n} < \infty.$$

4. if $\tau \leq 1$,
$$\sum_{\substack{p \text{ prime, } n \geq 2, \\ p^n \leq x}} f(p^n) \ll \frac{x}{\log x}.$$

Then
$$\sum_{n \leq x} f(n) = (1 + o(1)) \frac{x}{\log x} \frac{e^{-\gamma \tau}}{\Gamma(\tau)} \prod_{p \leq x} \left(1 + \frac{f(p)}{p} + \frac{f(p^2)}{p^2} + \dots \right).$$

Here γ is Euler's constant.

We now evaluate the counting function $\mathcal{Q}(\mathcal{T})(x)$ (which is quite standard, see also [8]). In fact, for our purposes it would basically suffice to have a reasonable lower bound for $\mathcal{Q}(\mathcal{T})(x)$, but since Wirsing's theorem readily supplies an asymptotic we will compute this.

Lemma 6.2.
$$\mathcal{Q}(\mathcal{T})(x) \sim C_{\mathcal{T}} \frac{x}{(\log x)^{1-\tau}},$$

where $C_{\mathcal{T}} > 0$ is a constant depending on $\mathcal{T}$.

Sketch proof of Lemma. The regularity condition on $\mathcal{T}$ in Theorem 2.4 guarantees that Wirsing's theorem on sums of multiplicative functions can be applied. We apply it to the multiplicative function $f(p^n) = \begin{cases} 1 & \text{if } p \in \mathcal{T} \\ 0 & \text{otherwise,} \end{cases}$
so that
$$\mathcal{Q}(\mathcal{T})(x) = \sum_{n \leq x} f(n).$$

Here
$$\sum_{p \leq x} \frac{f(p) \log p}{p} = \tau \log x + C + o(1),$$

and all further hypotheses of Wirsing's theorem are satisfied as well. Thus

$$
\begin{aligned}
\sum_{n \leq x} f(n) &= (1 + o(1)) \frac{x}{\log x} \frac{e^{-\gamma \tau}}{\Gamma(\tau)} \prod_{p \leq x} \left(1 + \frac{f(p)}{p} + \frac{f(p^2)}{p^2} + \dots \right) \\
&= (1 + o(1)) \frac{x}{\log x} \frac{e^{-\gamma \tau}}{\Gamma(\tau)} \exp \left(\sum_{p \leq x, p \in \mathcal{T}} \log(1 + \frac{1}{p-1}) \right) \\
&\sim C_{\mathcal{T}} \frac{x}{\log x} \exp(\tau \log \log x) \\
&\sim C_{\mathcal{T}} \frac{x}{(\log x)^{1-\tau}}.
\end{aligned}
$$

Here we used that $\sum_{p \leq x, p \in \mathcal{T}} \log(1 + \frac{1}{p-1}) = \sum_{p \leq x, p \in \mathcal{T}} 1/p + C'_{\mathcal{T}} + o(1)$, and that from $\sum_{p \leq x, p \in \mathcal{T}} \frac{\log p}{p} = \tau \log x + C + o(1)$ it follows by partial summation that $\sum_{p \leq x, p \in \mathcal{T}} 1/p = \tau \log \log x + C''_{\mathcal{T}} + o(1)$.

$\square$

Proof of Theorem 2.4. We intend to apply Theorem 4.1 (1), putting $\mathcal{S} = \mathcal{Q}(\mathcal{T}) \cap [1, x]$ and with $\mathcal{P}_0 = (\mathcal{P} \setminus \mathcal{T})$ as the set of sieving primes. In view of the foregoing calculation we have $\sigma = \mathcal{Q}(\mathcal{T})(x)/x = (1 + o(1))C_{\mathcal{T}} \log^{\tau-1} x$ in the theorem. Also, in view of the regularity condition on $\mathcal{T}$ (and partial summation) we have

$$\sum_{y/2 < p \leq y, p \in \mathcal{P}_0} \log p = (1 + o(1))\frac{y}{2} - \sum_{y/2 < p \leq y, p \in \mathcal{T}} \log p \; = \; (1 + o(1))(1 - \tau)\frac{y}{2}$$

$$= \; (1 + o(1))(1 - \tau) \sum_{y/2 < p \leq y} \log p.$$

Hence we choose $c = (1 + o(1))(1 - \tau)$ in Theorem 4.1.

If we set $\mathcal{P}_0^* = (\mathcal{P} \setminus \mathcal{T}) \cap [K^3, \infty)$, where $K = 1000(\sigma c^2)^{-1}(\log x)^2$; and we define a completely multiplicative function $f(n)$ by $f(p) = \begin{cases} 2 & \text{if } p \in \mathcal{P}_0^* \\ 0 & \text{otherwise} \end{cases}$; and we let $d(n)$ denote the divisor function; then

$$\sum_{\substack{1 < q \leq \sqrt{x}, \\ p|q \Rightarrow p \in \mathcal{P}_0^*}} \mu^2(q) \prod_{p|q} \frac{2}{p} \gg \sum_{q \leq \sqrt{x}} \frac{f(q)}{q} \; \gg \; \prod_{2 < p \leq \sqrt{x}} \left(1 - \frac{2}{p}\right) \prod_{p \leq \sqrt{x}} \left(1 - \frac{f(p)}{p}\right)^{-1} \sum_{n \leq \sqrt{x}} \frac{d(n)}{n}$$

$$\gg \; \prod_{p \leq \sqrt{x}} \left(1 - \frac{f(p)}{p}\right)^{-1}$$

$$\gg \; \exp\left(\sum_{K^3 \leq p \leq \sqrt{x}, p \in \mathcal{P} \setminus \mathcal{T}} \frac{2}{p}\right)$$

$$\gg \; ((\log x)/\log \log x)^{2 - 2\tau}.$$

In the first inequality here we added the non-squarefree values of q to the sum, (which, as the reader may easily convince themselves, does not change the order of magnitude), and in the second inequality we used the general lower bound (**), from the proof of Proposition 4.2.

We conclude that the left hand side is much larger than $\sigma^{-1} \asymp \log^{1-\tau} x$, and so the "Sieve Controls Size" condition in Theorem 4.1 is satisfied. Applying that theorem, we immediately obtain the upper bound

$$\max(\mathcal{A}(x), \mathcal{B}(x)) \ll_{\mathcal{T}} \sqrt{x}(\log x)^4.$$

$\square$

24

Proof of Corollary 2.5. The proof of the corollary is as before. Suppose that $\mathcal{A} + \mathcal{B} + \mathcal{C} \sim \mathcal{Q}(\mathcal{T})$. Then by the binary result $(\mathcal{A} + \mathcal{B})(x), (\mathcal{A} + \mathcal{C})(x), (\mathcal{B} + \mathcal{C})(x) \sim x^{1/2 + o(1)}$, and so by Ruzsa's inequality (Lemma 3.5)

$$\frac{x^2}{(\log x)^{2-2\tau}} \ll_{\mathcal{T}} \mathcal{Q}(\mathcal{T})(x)^2 \ll ((\mathcal{A} + \mathcal{B} + \mathcal{C})(x))^2 \leq x^{3/2 + o(1)},$$

a contradiction. $\qquad\square$

7 Proof of Theorem 2.6, Ostmann's problem revisited

In this section we shall prove Theorem 2.6. In view of the existing results on Ostmann's problem (see [5], for example), we may assume that in a putative decomposition of the primes we have, for given x,

$$\frac{\sqrt{x}}{\log^5 x} \leq \mathcal{A}(x) \leq \mathcal{B}(x),$$

say. We will show that, in order for the lower bound on $\mathcal{A}(x)$ not to be violated, we must have $\mathcal{B}(x) \ll \sqrt{x} \log\log x$, which (since we must have $\mathcal{A}(x)\mathcal{B}(x) \gg \pi(x)$) immediately implies the claimed result that

$$\frac{\sqrt{x}}{\log x \log\log x} \ll \mathcal{A}(x) \leq \mathcal{B}(x) \ll \sqrt{x} \log\log x.$$

We will do this by performing a bit more delicate analysis, requiring a lower bound estimate of Hildebrand [20], of the final applications of the large and larger sieves in our general proofs.

Suppose, then, that for each prime $p \leq Y := \sqrt{x}/\log^{10} x$ the set $\mathcal{A}$ occupies $\nu_{\mathcal{A}}(p) = p/2 + \epsilon_p$ residue classes modulo p. This leaves $\nu_{\mathcal{B}}(p) \leq p/2 - \epsilon_p$ residue classes for the set $\mathcal{B}$ to occupy, since none of the sums $a + b$ may be divisible by p. Now

$$
\begin{aligned}
\sum_{p \leq Y} \frac{\log p}{p/2 + \epsilon_p} + \sum_{p \leq Y} \frac{\log p}{p/2 - \epsilon_p} &= \sum_{p \leq Y} \frac{p \log p}{(p/2)^2 - \epsilon_p^2} \\
&= 4 \sum_{p \leq Y} \frac{\log p}{p} + 4 \sum_{p \leq Y} \frac{\log p}{p} \cdot \frac{\epsilon_p^2}{(p/2)^2 - \epsilon_p^2} \\
&\geq 2 \log x - 40 \log\log x + 16 \sum_{p \leq Y} \frac{\log p}{p} \frac{\epsilon_p^2}{p^2} + O(1),
\end{aligned}
$$

since $\sum_{p \leq Y} (\log p)/p = \log Y + O(1) = (1/2)\log x - 10 \log\log x + O(1)$. But the left hand side cannot exceed $2(\log x + 1)$, because in that case we would have either

$$\sum_{p \leq Y} \frac{\log p}{\nu_{\mathcal{A}}(p)} \geq \log x + 1 \qquad \text{or} \qquad \sum_{p \leq Y} \frac{\log p}{\nu_{\mathcal{B}}(p)} \geq \log x + 1,$$

and then by the larger sieve (that is Lemma 3.3, with $\mathcal{P}_0$ chosen as $\{p \leq Y\}$) we would have either $\mathcal{A}(x) \ll Y$ or $\mathcal{B}(x) \ll Y$, which we know to be false. Therefore we must have the following bound:

$$\sum_{p \leq Y} \frac{\log p}{p} \frac{\epsilon_p^2}{p^2} \ll \log \log x.$$

We shall use the bound that we just derived to estimate the denominator in a final application of the large sieve. Firstly we note that, by the Cauchy–Schwarz inequality and standard results on the distribution of prime numbers,

$$\sum_{\log x \leq p \leq Y} \frac{1}{p} \frac{|\epsilon_p|}{p} \leq \sqrt{\sum_{\log x \leq p \leq Y} \frac{1}{p \log p} \cdot \sum_{\log x \leq p \leq Y} \frac{\log p}{p} \frac{\epsilon_p^2}{p^2}} \ll \sqrt{\frac{1}{\log \log x} \cdot \log \log x} \ll 1.$$

In particular, this implies that $\sum_{\log x \leq p \leq Y; |\epsilon_p| \geq p/4} 1/p \ll 1$, and since we have $\sum_{Y < p \leq \sqrt{x}} 1/p \ll 1$ we clearly also have

$$\sum_{\log x \leq p \leq \sqrt{x}} \frac{1}{p} \frac{|\epsilon_p|}{p} \ll 1 \quad \text{and} \quad \sum_{\log x \leq p \leq \sqrt{x}; |\epsilon_p| \geq p/4} \frac{1}{p} \ll 1.$$

Now if we define a multiplicative function $f(n)$, supported on squarefree integers, by setting

$$f(p) := \begin{cases} \frac{\nu_{\mathcal{A}}(p)}{p - \nu_{\mathcal{A}}(p)} = \frac{p/2 + \epsilon_p}{p/2 - \epsilon_p} & \text{if } |\epsilon_p| \leq p/4 \\ 0 & \text{otherwise} \end{cases},$$

then the large sieve (that is Lemma 3.2, with the choice $Q = \sqrt{x}$) implies

$$\mathcal{B}(x) \leq \frac{2x}{\sum_{n \leq \sqrt{x}} f(n)}.$$

It remains to show that $\sum_{n \leq \sqrt{x}} f(n) \gg \sqrt{x}/\log \log x$, and to do this we shall use the following nice result, which is a slightly simplified version of Theorem 2 of Hildebrand [20].

Theorem 7.1 (Hildebrand, 1987). *Let $X \geq Z \geq 2$. Let f be a multiplicative function supported on squarefree numbers, and satisfying $0 \leq f(p) \leq K$ for some constant $K \geq 1$, for all primes p. Then*

$$\frac{1}{X} \sum_{n \leq X} f(n) \geq \frac{e^{-\gamma(K-1)}}{\Gamma(K)} \prod_{p \leq X} \left(1 - \frac{1}{p}\right) \left(1 + \frac{f(p)}{p}\right) \cdot$$

$$\cdot \left(\sigma_- \left(e^{\sum_{Z \leq p \leq X}(1 - f(p))^+/p}\right) \left(1 + O\left(\left(\frac{\log Z}{\log X}\right)^{\alpha}\right)\right) + O(e^{-((\log X)/(\log Z))^{\alpha}})\right),$$

where $\gamma \approx 0.577$ is Euler's constant; $(1 - f(p))^+$ denotes the positive part of $(1 - f(p))$; $\alpha > 0$ is an absolute constant; the implicit constant in the "big Oh" terms depends on K only; and $\sigma_-(u)$ is a certain fixed function of $u \geq 1$ that satisfies $\sigma_-(u) \gg u^{-u}$.

We can apply the theorem to the function $f(n)$ that we defined above, with the choices $X = \sqrt{x}$, $Z = \log x$ and $K = 3$. We find that

$$\sum_{\log x \leq p \leq \sqrt{x}} \frac{(1 - f(p))^+}{p} \ll \sum_{\substack{\log x \leq p \leq \sqrt{x}, \\ |\epsilon_p| > p/4}} \frac{1}{p} + \sum_{\log x \leq p \leq \sqrt{x}} \frac{|\epsilon_p|}{p^2} \ll 1,$$

in view of our previous calculations, so that

$$\sum_{n \leq \sqrt{x}} f(n) \gg \sqrt{x} \prod_{p \leq \sqrt{x}} \left(1 - \frac{1}{p}\right)\left(1 + \frac{f(p)}{p}\right) \gg \sqrt{x} \prod_{p \leq \log x} \left(1 - \frac{1}{p}\right) \gg \frac{\sqrt{x}}{\log \log x},$$

as claimed. $\qquad\qquad\square$

References

[1] N. Alon, A. Granville and A. Ubis, The number of sumsets in a finite field, *Bull. Lond. Math. Soc.* **42** (2010), 784–794.

[2] E. S. Croot, C. Elsholtz, On thin sets of primes expressible as sumsets, *Acta Math. Hungar.* **106** (2005), 197–226.

[3] C. Dartyge, A. Sárközy, On additive decompositions of the set of primitive roots modulo p, *Monatsh. Math.* **169** (2013), no. 3-4, 317–328.

[4] C. Elsholtz, A Remark on Hofmann and Wolke's Additive Decompositions of the Set of Primes, *Arch. Math.* **76** (2001), 30–33.

[5] C. Elsholtz, The Inverse Goldbach Problem, *Mathematika* **48** (2001), 151–158.

[6] C. Elsholtz, Some remarks on the additive structure of the set of primes, Number theory for the millennium, I (Urbana, IL, 2000), 419–427, (Proceedings of the Millennial Conference on Number Theory (Bennett et.al.), AK Peters, 2002).

[7] C. Elsholtz, Combinatorial Prime Number Theory, Habilitationsschrift, TU Clausthal, 2002.

[8] C. Elsholtz, Additive decomposability of multiplicatively defined sets, *Funct. Approx. Comment. Math.* **35** (2006), 61–77.

[9] C. Elsholtz, Multiplicative decomposability of shifted sets, *Bull. London Math. Soc.* **40** (2008), 97–107.

[10] P. Erdős, Problems in number theory and combinatorics, Proceedings of the Sixth Manitoba Conference on Numerical Mathematics (Univ. Manitoba, Winnipeg, Man., 1976), Congress. Numer. XVIII, pp. 35–58, Utilitas Math., Winnipeg, Man.

[11] P. Erdős, Some problems on number theory, Analytic and elementary number theory (Marseille, 1983), Publ. Math. Orsay, 86-1, pp. 53–67, Univ. Paris XI, Orsay, 1986.

[12] P. Erdős, M.B. Nathanson, Sets of natural numbers with no minimal asymptotic bases, *Proc. Amer. Math. Soc.* **70** (1978), 100–102.

[13] E. Fouvry, G. Tenenbaum, Répartition statistique des entiers sans grand facteur premier dans les progressions arithmétiques, *Proc. London Math. Soc.* (3) **72** (1996), no. 3, 481–514.

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

[15] P. X. Gallagher, A larger sieve, *Acta Arith.* **18** (1971), 77–81.

[16] A. Granville, K. Soundararajan, *Multiplicative number theory*, course notes, 2011.

[17] B. J. Green, A. J. Harper, Inverse questions for the large sieve, *Preprint*, arXiv:1311.6176.

[18] K. Gyarmati, S. Konyagin, A. Sárközy, On the reducibility of large sets of residues modulo p, *J. Number Theory* **133** (2013), no. 7, 2374–2397.

[19] A. J. Harper, Bombieri–Vinogradov and Barban–Davenport–Halberstam type theorems for smooth numbers, *Preprint*, arXiv:1208.5992.

[20] A. Hildebrand, Quantitative mean value theorems for nonnegative multiplicative functions. II, *Acta Arith.* **48** (1987), no. 3, 209–260.

[21] A. Hofmann, D. Wolke, On additive decompositions of the set of primes, *Arch. Math.* **67** (1996), 379–382.

[22] B. Hornfeck, Ein Satz über die Primzahlmenge, *Math. Z.* **60** (1954), 271–273, see also the Zentralblatt review by Erdős and the correction in *Math. Z.* **62** (1955), page 502.

[23] W. B. Laffer, H. B. Mann, Decomposition of sets of group elements, *Pacific J. Math.* **14** (1964) 547–558.

[24] H. B. Mann, *Addition theorems: The addition theorems of group theory and number theory*, Wiley Interscience, New York-London-Sydney, 1965.

[25] H. Montgomery, The analytic principle of the large sieve, *Bull. Amer. Math Soc.* **84** (1978), 547–567.

[26] H. L. Montgomery, R. C. Vaughan, *Multiplicative Number Theory I: Classical Theory,* first edition, published by Cambridge University Press, 2007.

[27] H.-H. Ostmann, *Additive Zahlentheorie*, 2 volumes, Springer-Verlag, Berlin-Heidelberg-New York, 1956, reprint 1968.

[28] C. Pomerance, A. Sárközy, C. L. Stewart, On Divisors of Sums of Integers III, *Pacific J. Math.* **133** (1988), 363–379.

[29] J.-C. Puchta, On additive decompositions of the set of primes, *Arch. Math.* **78** (2002), 24–25.

[30] I.Z. Ruzsa, Sumsets and structure. Combinatorial number theory and additive group theory, 87–210, Adv. Courses Math. CRM Barcelona, Birkhäuser Verlag, Basel, 2009.

[31] A. Sárközy, Über totalprimitive Folgen, *Acta Arith.* **8** (1962/1963), 21–31.

[32] A. Sárközy, Über reduzible Folgen, *Acta Arith.* **10** (1964/1965), 399–408.

[33] A. Sárközy, Unsolved problems in number theory, *Period. Math. Hungar.* **42** (2001), no. 1-2, 17–35.

[34] A. Sárközy, On additive decompositions of the set of quadratic residues modulo p, *Acta Arith.* **155** (2012), no. 1, 41–51.

[35] A. Sárközy and E. Szemerédi, On the sequence of squares (in Hungarian), *Mat. Lapok* **16** (1-2) (1965), 76-85.

[36] I. D. Shkredov, Sumsets in quadratic residues, *Acta Arith.*, (to appear), arXiv:1305.4093

[37] I. Shparlinski, Additive Decompositions of Subgroups of Finite Fields, *SIAM J. Discrete Math.* **27** (2013), no. 4, 1870–1879.

[38] R. Tijdeman, On integers with many small prime factors. *Compositio Math.* **26** (1973), 319–330.

[39] E. Wirsing, Ein metrischer Satz über Mengen ganzer Zahlen, *Arch. Math.* **4** (1953), 392–398.

[40] E. Wirsing, Das asymptotische Verhalten von Summen über multiplikative Funktionen. II, *Acta Math. Acad. Sci. Hungar.* **18** (1967), 411–467.

[41] E. Wirsing, *Über additive Zerlegungen der Primzahlmenge*, lecture at Oberwolfach, an abstract can be found in Tagungsbericht 28/1972. An unpublished written version (1998) exists with the title *On the additive decomposability of the set of primes, A Memo.*

Christian Elsholtz, Institut für Mathematik A, Technische Universität Graz, Steyrergasse 30/II, A-8010 Graz, Austria
email: elsholtz@math.tugraz.at

Adam J. Harper, Department of Pure Mathematics and Mathematical Statistics, Wilberforce Road, Cambridge CB3 0WA, England
email: A.J.Harper@dpmms.cam.ac.uk