ACTA ARITHMETICA
Online First version

Growth rates of sequences governed by the squarefree properties of their translates

by

Wouter van Doorn and Terence Tao

Abstract. We answer several questions of Erdős regarding sequences of natural numbers A whose translates $n + A$ intersect with the squarefree numbers in various specified ways. For instance, we show that if every translate only contains finitely many squarefree numbers, then A has zero density, although the decay rate of this density can be arbitrarily slow. On the other hand, there exist sequences A with optimal density $6/\pi^2$ for which infinitely many n exist such that $n + a$ is squarefree for all $a \in A$ with $a < n$. In fact, infinitely many such n exist for every exponentially increasing sequence, as long as the sequence avoids at least one residue class modulo p^2 for all primes p, a property we call admissible. If one instead requires infinitely many n to exist such that $n + a$ is squarefree for all $a \in A$, then A can have density arbitrarily close to, but not equal to, $6/\pi^2$. Finally, we prove bounds on the largest admissible subset of $\{1, \dots, N\}$.

1. Introduction. In 1981, Erdős [Er81] stated various unconventional questions on additive and multiplicative number theory. In particular, he asked about the growth rates of sequences of positive integers whose translates are constrained by various properties concerning the squarefree natural numbers

$$\mathcal{SF} \coloneqq \mathbb{N} \setminus \bigcup_p p^2\mathbb{N} = \{1, 2, 3, 5, 6, 7, 10, \dots\} \quad \text{(OEIS A005117)}.$$

As is well known, by Euler's solution to the Basel problem this set has natural density

$$\prod_p \left(1 - \frac{1}{p^2}\right) = \frac{1}{\zeta(2)} = \frac{6}{\pi^2} \quad \text{(OEIS A059956)}, \tag{1.1}$$

2020 *Mathematics Subject Classification*: Primary 11B05; Secondary 11B50, 11N25.
Key words and phrases: squarefree integer, translate, dense sequence.
Received 7 December 2025.
Published online 10 July 2026.

DOI: 10.4064/aa251207-28-5 © Instytut Matematyczny PAN, 2026

and it is not hard to show ([1])

$$|\mathcal{SF} \cap [x]| = \frac{6}{\pi^2}x + O(\sqrt{x}) \quad \text{(OEIS A013928)}.$$

In fact, this error term can be improved, and one has

$$|\mathcal{SF} \cap [x]| = \frac{6}{\pi^2}x + O\left(\frac{\sqrt{x}}{\exp(c\log^{3/5} x/(\log\log x)^{1/5})}\right) \tag{1.2}$$

for all $x \geq 3$ and some absolute constant $c > 0$; see [Wa63]. One can improve the error term further if one assumes the Riemann hypothesis: see [Li16] (or [Bl, Problem 969]) for the most recent results in this direction.

Now, for each positive integer n, the translate $\mathcal{SF} - n$ denotes those integers a for which $n+a$ is squarefree. We recall the following four definitions from [Er81], and add three more that are implicitly used there.

Definition 1.1 (Properties P, Q, $\overline{P}$, $\overline{P}_\infty$, etc.). Let A be a set or sequence of natural numbers. If we consider A (which can be finite or infinite) to be a sequence, we always assume it is strictly increasing.

- We say that A has *property* P if $\mathcal{SF} - n$ has finite intersection with A for every $n \in \mathbb{N}$. That is, for all n, one has $n + a$ squarefree for only finitely many $a \in A$.
- We say that A has *property* Q if $\mathcal{SF} - n$ contains $A \cap [n-1]$ for infinitely many $n \in \mathbb{N}$. That is, for infinitely many n, one has $n + a$ squarefree for all $a \in A$ with $a < n$.
- We say that A has *property* $\overline{P}$ if $\mathcal{SF} - n$ contains A for infinitely many $n \in \mathbb{N}$. That is, for infinitely many n, one has $n + a$ squarefree for all $a \in A$.
- We say that A has *property* $\overline{P}_\infty$ if $A \setminus (\mathcal{SF} - n)$ is finite for infinitely many $n \in \mathbb{N}$. That is, for infinitely many n, one has $n + a$ squarefree for all but finitely many $a \in A$.
- We say that A has *squarefree sums* if $\mathcal{SF} - a$ contains A for all $a \in A$. That is, $a + a' \in \mathcal{SF}$ for all $a, a' \in A$.
- We say that A is *admissible* if A avoids at least one residue class $b_p \pmod{p^2}$ for each prime p.
- We say that A is *almost admissible* if for each prime p there is a residue class $b_p \pmod{p^2}$ that intersects only finitely many elements of A.

We remark that all of these properties are downward monotone: whenever a set A has any of the above properties, then so does any subset of A. Property P, being additionally closed under finite unions, in fact forms an order ideal. Furthermore, we have the following easy implications (summarized in Figures 1, 2), most of which are already implicit in [Er81]:

([1]) See Section 1.8 for our notation conventions.

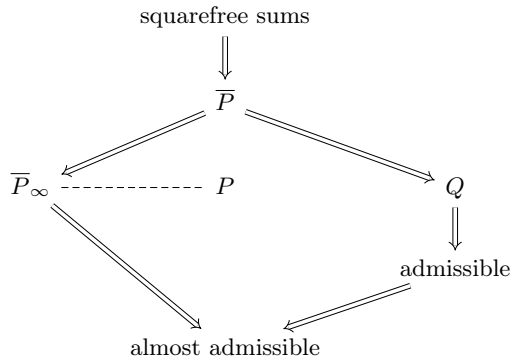

Fig. 1. Implications between various properties for infinite sets A. The dashed line indicates the incompatibility of P and $\overline{P}_\infty$ in the infinite case. Note how P stands in opposition to the other six properties, which are tied to each other by various implications.

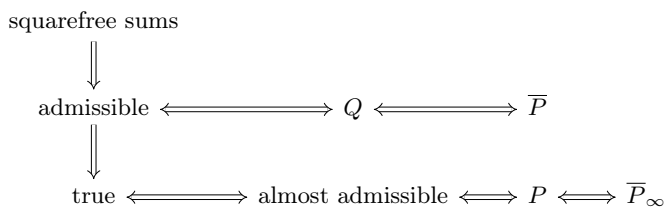

Fig. 2. Implications and equivalences between various properties for finite sets A.

- Every infinite set with squarefree sums has property $\overline{P}$.
- Every set with property $\overline{P}$ has properties $\overline{P}_\infty$ and Q as well.
- No infinite set with property P obeys $\overline{P}_\infty$ (let alone $\overline{P}$).
- Every admissible set is almost admissible.
- Every set with property Q is admissible.
- Every set with property $\overline{P}_\infty$ is almost admissible.
- Properties P and $\overline{P}_\infty$, as well as the almost admissible property, are unaffected by the addition or deletion of a finite number of elements. In particular, finite sets obey all three properties.
- Finite sets obey $\overline{P}$ (hence Q) if and only if they are admissible.

Erdős also listed the specific sequences

$$A_1 \coloneqq \{2^j + 1 : j \in \mathbb{N}\} = \{3, 5, 9, 17, 33, \dots\} \qquad \text{(OEIS A000051)}$$
$$A_2 \coloneqq \{2^j - 1 : j \in \mathbb{N}\} = \{1, 3, 7, 15, 31, \dots\} \qquad \text{(OEIS A000225)}$$
$$A_3 \coloneqq \{j! + 1 : j \in \mathbb{N}\} = \{2, 3, 7, 25, 121, \dots\} \qquad \text{(OEIS A038507)}$$
$$A_4 \coloneqq \{j! - 1 : j > 1\} = \{1, 5, 23, 119, 719, \dots\} \qquad \text{(OEIS A033312)}$$

as test cases for the above properties.

1.1. Property P. Regarding property P, Erdős ([Er81, p. 179], [Bl, Problem 1102]) writes

> Probably a sequence having property P must increase fairly fast, but I have no results in this direction.

Contrary to Erdős' expectations, sequences with property P do not necessarily have to grow very quickly. In Section 3 we show

THEOREM 1.2.

(i) *If $A = \{a_1 < a_2 < \cdots\}$ has property P, then A has natural density 0. That is, the ratio a_j/j goes to infinity with j.*
(ii) *Conversely, for any function $f\colon \mathbb{N} \to \mathbb{N}$ that goes to infinity, there exists an infinite sequence $\{a_1 < a_2 < \cdots\}$ with property P for which $a_j/j \le f(j)$ holds for all $j \in \mathbb{N}$.*

1.2. Property Q. By standard sieve theory and (1.1), every admissible (or almost admissible) sequence has upper density at most $6/\pi^2$. Since sequences with property Q are automatically admissible, we conclude

THEOREM 1.3 (Upper bound). *Every sequence with property Q has upper density at most $6/\pi^2$.*

Erdős ([Er81, p. 179], [Bl, Problem 1102]) writes

> It is easy to see that if A increases sufficiently fast [and is admissible] then it has property Q and in fact there is an n, $a_k < n < a_{k+1}$ ([2]) for which $n + a_i$, $i = 1, \ldots, k$, is always squarefree. I have no precise information about the rate of increase a sequence having property Q must have.

Perhaps surprisingly, the bound in Theorem 1.3 is sharp; in Section 4.1 we will show

THEOREM 1.4 (Lower bound). *There exists an infinite sequence $A = \{a_1 < a_2 < \cdots\} \subset \mathcal{SF}$ with property Q which has natural density $6/\pi^2$. Equivalently, $a_j/j \to \pi^2/6$ as $j \to \infty$.*

In Section 4.2 we also quantify what counts as "sufficiently fast" in Erdős' claim:

THEOREM 1.5 (Sufficiently fast sequences). *There is an absolute constant C such that, if $A = \{a_1 < a_2 < \cdots\}$ is an admissible sequence with $a_j \ge \exp(Cj/\log j)$ for infinitely many j, then A has property Q.*

In fact, as we will explain in Section 4.3, one can take C to be any constant larger than 4. If one then furthermore adds a regularity condition on the growth of A, the further claim by Erdős that suitable n exist between consecutive elements of A also follows.

([2]) Erdős actually writes n_{7k+1} instead of a_{k+1}, but we assume that this is simply a typographical error.

THEOREM 1.6 (Sufficiently fast sequences, alternate version). *If* $A = \{a_1 < a_2 < \cdots\}$ *is an admissible sequence with* $a_j \geq \max(\exp(5j/\log j), a_{j-1} + a_{j-1}^{10/11})$ *for all sufficiently large* j*, then for all sufficiently large* j *there exists an* $n \in \mathbb{N}$ *with* $a_{j-1} < n < a_j$ *such that* $n + a_i \in \mathcal{SF}$ *for all* $i < j$*.*

In the converse direction, we can find rapidly growing admissible sequences that do *not* obey property Q:

THEOREM 1.7 (Fast growing counterexample). *There exists an absolute constant* $c > 0$ *and an admissible sequence* $A = \{a_1 < a_2 < \cdots\} \subset \mathcal{SF}$ *which does not obey property* Q*, and for which the inequality* $a_j \geq \exp(cj^{1/2}/\log^{1/2} j)$ *holds for all sufficiently large* j*.*

We establish this result in Section 4.4.

1.3. Squarefree sums. As observed in [Er81, p. 179], a straightforward greedy construction gives a sequence $A = \{a_1 < a_2 < \cdots\}$ with squarefree sums. Erdős writes

> In fact one can find a sequence which grows exponentially. Must such a sequence really increase so fast? I do not expect that there is such a sequence of polynomial growth.

Let ES_N denote the size of the largest subset of $[N]$ with squarefree sums. Erdős and Sárközy [ES87], Sárközy [Sa92], Gyarmati [Gy01] and Konyagin [Ko04] have given upper and lower bounds on this quantity, with the current best bounds of

$$\log^2 N \log\log N \ll \mathrm{ES}_N \ll N^{11/15} \exp\left(O\left(\frac{\log N}{\sqrt{\log\log N}}\right)\right)$$

for all large N established in [Ko04]. (For comparison, the large sieve inequality in Theorem A.2 only gives the weaker upper bound $\mathrm{ES}_N \ll N^{3/4}$.) Inverting the upper bound, we obtain a lower bound

$$a_j \gg j^{15/11} \exp\left(-O\left(\frac{\log j}{\sqrt{\log\log j}}\right)\right)$$

for all large j and any infinite sequence $A = \{a_1 < a_2 < \cdots\}$ with squarefree sums. It is likely that the constructions in [Ko04] can be modified to generate an infinite sequence $A = \{a_1 < a_2 < \cdots\}$ with squarefree sums such that

$$\log^2 N \log\log N \ll |A \cap [N]|$$

for all large N, although we will not do so here. Assuming that this is the case, such a sequence would obey the upper bound

$$a_j \ll \exp(O(j^{1/2}/\log^{1/2} j)).$$

We remark that a modification of the construction in Theorem 1.5 gives an infinite sequence $A = \{a_1 < a_2 < \cdots\}$ for which the weaker upper bound

$$a_j \ll \exp(O(j/\log j))$$

holds. We refer the reader to the first version of this paper ([3]) for details.

We also note that the opposite problem of studying sets with very few (or no) squarefree pairwise sums was considered in [Na89], [Fi93], [LNS90], [Sc05].

REMARK 1.8. Erdős [Er81, pp. 179–180] also writes

> Is there a sequence of integers $1 \le a_1 < a_2 < \cdots$ so that for every i, $a_i \equiv t \pmod{p^2}$ implies $1 \le t < p^2/2$? If such a sequence exists then clearly $a_i + a_j$ is always squarefree, but I am doubtful if such a sequence exists. I formulated this problem while writing these lines and must ask the indulgence of the reader if it turns out to be trivial.

Indeed, as noted in [Bl, Problem 1103], it is easy to rule out such a sequence: by the prime number theorem, for i large enough there must exist a prime p between $\sqrt{a_i}$ and $\sqrt{2a_i}$, and it is clear for such a prime that the residue $t = a_i$ of $a_i \pmod{p^2}$ will be larger than $p^2/2$.

1.4. Properties $\overline{P}$ and $\overline{P}_\infty$. In contrast to the previous results on squarefree sums, sequences with properties $\overline{P}$ and $\overline{P}_\infty$ can be far denser:

THEOREM 1.9.

(i) *If A has property $\overline{P}$ or $\overline{P}_\infty$, then the upper density of A is strictly less than $6/\pi^2$.*

(ii) *Conversely, for any $\varepsilon > 0$, one can find a sequence A obeying property $\overline{P}$ (and hence $\overline{P}_\infty$) whose lower density (and hence upper density) is at least $6/\pi^2 - \varepsilon$.*

We prove this result in Section 5. We summarize some of the above results in Table 1.

Table 1. The admissibility and possible densities of (infinite) sequences with the various properties identified by Erdős

Property	Admissibility	Density
P	arbitrary	0 (arbitrarily slowly)
Q	admissible	$\le \frac{6}{\pi^2}$ (equality attainable)
$\overline{P}$	admissible	$< \frac{6}{\pi^2}$ (equality almost attainable)
$\overline{P}_\infty$	almost admissible	$< \frac{6}{\pi^2}$ (equality almost attainable)
squarefree sums	admissible	0 (with at least $O(x^{-1/4})$ decay)

1.5. The size of admissible sets. In [Er81, p. 180], Erdős defines $A(x)$ (OEIS A083544) to be the maximal value of $|A \cap [x]|$, where A ranges over

([3]) arXiv:2512.01087v1.

all admissible sets. By the Chinese remainder theorem and standard sieve theory, $A(x)$ is also the maximal value of $\sum_{y\le n<y+x}\mu^2(n)$, where μ is the Möbius function and where y ranges over all positive integers. From (1.2) one has the lower bound

$$A(x) \ge |\mathcal{SF}\cap[x]| = \frac{6}{\pi^2}x + O\left(\frac{\sqrt{x}}{\exp(c\log^{3/5}x/(\log\log x)^{1/5})}\right)$$

for $x \ge 3$. Erdős [Er81, p. 180] states that Ruzsa proved that the strict inequality $A(x) > |\mathcal{SF}\cap[x]|$ holds for infinitely many x and writes

> Probably this holds for all large x. It would be of some interest to estimate $A(x)$ as accurately as possible.

By improving upon the aforementioned lower bound on $A(x)$, we can confirm Erdős' conjecture.

THEOREM 1.10. *For sufficiently large x, one has*

$$\frac{\sqrt{x}}{\log x} \ll A(x) - \frac{6}{\pi^2}x \ll x^{4/5}.$$

In particular, $A(x) > |\mathcal{SF}\cap[x]|$ for all large x.

We establish this result in Section 6. By comparing the OEIS sequences A013928 and A083544 (see Figure 3) it in fact seems likely that $A(x) > |\mathcal{SF}\cap[x]|$ holds for all $x \ge 18$.

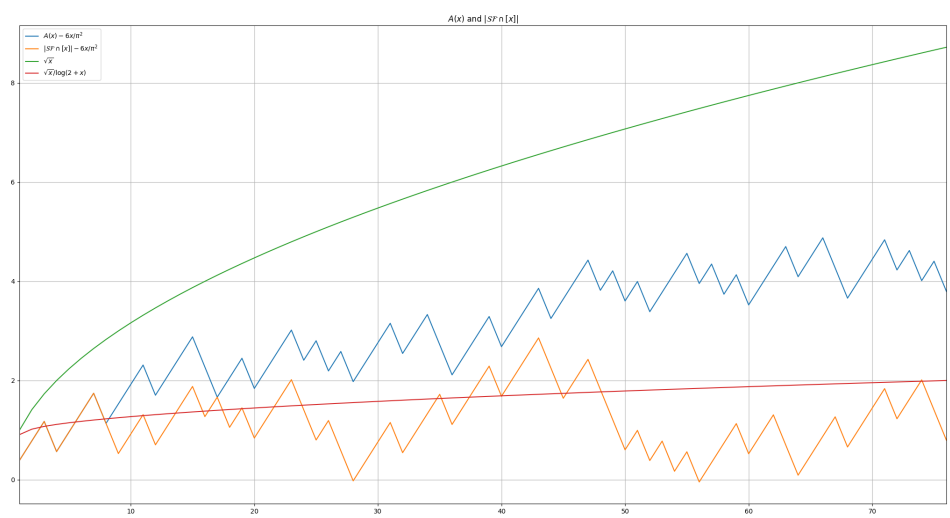

Fig. 3. A comparison of $A(x)$ (OEIS A083544) and $|\mathcal{SF}\cap[x]|$ (OEIS A013928), after subtracting off the main term $\frac{6}{\pi^2}x$. All data is drawn from the OEIS. The numerical range is too short to reach any definitive prediction about the growth rate of $A(x) - \frac{6}{\pi^2}x$, although the upper bound of $O(x^{4/5})$ is likely quite far from the truth.

1.6. Specific sequences. We can specialize the above discussion to the four specific sequences A_1, A_2, A_3, A_4 singled out in [Er81]. We first of all claim that all four sequences are admissible (and thus almost admissible). Indeed, A_1 avoids 0 (mod 4) and 1 (mod p^2) for all odd primes p, and A_2 similarly avoids 2 (mod 4) and -1 (mod p^2) for all odd primes p. Further, both A_3 and A_4 avoid 0 (mod 4), while the elements $j! \pm 1$ are congruent to ± 1 (mod p^2) for all $j \geq 2p$. Thus, for odd p, A_3 and A_4 occupy at most $2p < p^2$ residue classes modulo p^2 and must therefore avoid at least one, demonstrating admissibility.

Since all four sequences grow faster than $\exp(Cj/\log j)$, we conclude that A_1, A_2, A_3, A_4 all obey property Q by Theorem 1.5. In fact, the slightly stronger conclusion of Theorem 1.6 also holds for these sequences. On the other hand, the specific sums $3+5$, $1+3$, $3+25$, and $1+23$ are all divisible by 2^2, showing that none of A_1, A_2, A_3, A_4 have squarefree sums.

As for properties $\overline{P}$ and $\overline{P}_\infty$, Artin's primitive root conjecture implies that 2 is a primitive root of $\mathbb{Z}/p\mathbb{Z}$ for infinitely many primes p (OEIS A001122). Under the additional assumption of the plausible conjecture that infinitely many of these primes are not Wieferich primes ([4]) (i.e., $p^2 \nmid 2^{p-1} - 1$), we conclude that 2 is a primitive root of $\mathbb{Z}/p^2\mathbb{Z}$ for infinitely many p. Assuming the latter, let $n \in \mathbb{N}$ be given, and let $p > n+1$ be a prime for which 2 is a primitive root of $\mathbb{Z}/p^2\mathbb{Z}$. We then obtain infinitely many a in both A_1 and A_2 for which $a+n$ is divisible by p^2, so that A_1, A_2 do not obey properties $\overline{P}$ or $\overline{P}_\infty$. It looks difficult to establish this fact unconditionally, however. For A_3, A_4, the situation seems delicate: for fixed squarefree n, the expression $j! + n$ is, for all $j \geq 2n$, not divisible by p^2 for any $p \leq j$, so by (2.1) below one heuristically predicts this number to be squarefree with a "probability" of $1 - O(1/(j \log j))$. The sum $\sum_j \frac{1}{j \log j}$ is barely divergent, suggesting that there are infinitely many j for which $j! + n$ is not squarefree, implying that A_3 and A_4 do not obey the properties $\overline{P}, \overline{P}_\infty$. Given the special structure of the sequence $j! + n$ however, we do not find this to be a completely convincing argument.

Probabilistic heuristics also suggest that A_1, A_2, A_3, A_4 do not have property P, but this looks difficult to establish with known techniques. Even just verifying that (say) $2^j + 1$ can be squarefree for infinitely many j, while extremely plausible from such heuristics, does not seem amenable to existing tools such as sieve theory due to the rapid growth of the sequence, even if one assumes powerful conjectures such as the abc conjecture or Artin's conjecture. In [Er81, p. 178] it is written that "of course it seems hopeless" to solve these problems.

([4]) While it is believed that infinitely many Wieferich primes exist, they are exceedingly rare in practice; the only known examples are 1093 and 3511 (OEIS A001220).

1.7. Primes and k-free sequences. Erdős [Er81, p. 179] also writes

> These problems can of course be stated for other sequences than p^2, but we formulate only one such question: Is there an infinite sequence $a_1 < a_2 < \cdots$ so that there are infinitely many n for which for all $a_k < n$, $n + a_k$ is always a prime?

It turns out that with the recent result [TZ23], this question is answered in the affirmative. Indeed, from [TZ23, Corollary 1.6] one can find $a_1 < a_2 < \cdots$ and $b_1 < b_2 < \cdots$ such that $a_i + b_j$ is prime for all $1 \leq i < j$. Taking a sufficiently sparse subsequence a_{j_l} for which $a_{j_{l+1}} > b_{j_l}$ holds for all l, then gives the claim.

Moving on to higher powers instead, without too much difficulty one can extend the results in this paper from squarefree integers to k-free integers, for any fixed $k \geq 2$. That is, by making the obvious necessary numerical and definitional changes, we may formulate the results for $\mathbb{N} \setminus \bigcup_p p^k\mathbb{N}$ as opposed to $\mathcal{SF}$. For instance, $\zeta(2) = \pi^2/6$ should then be replaced by $\zeta(k)$ everywhere, suitable numerical adjustments should be made to Theorem 1.6, the exponents $1/2$ in Theorem 1.7 are to be replaced by $1/k$ (and $\sqrt{x}$ similarly replaced by $x^{1/k}$ in (1.2) and Theorem 1.10), and the exponent $4/5$ in Theorem 1.10 should be replaced by $\frac{2k}{3k-1}$. We leave the full proofs of these generalizations to the interested reader.

1.8. Notation. All sums and products over p are understood to be over primes, with $2 = p_1 < p_2 < \cdots$ denoting the sequence of primes. If E is a finite set, we use $|E|$ to denote its cardinality. For any x, we use $[x]$ to denote the discrete interval

$$[x] \coloneqq \{n \in \mathbb{N} : n \leq x\}.$$

If A is a set or sequence of natural numbers, we define its upper and lower density to be

$$\limsup_{x\to\infty} \frac{|A \cap [x]|}{x} \quad \text{and} \quad \liminf_{x\to\infty} \frac{|A \cap [x]|}{x}$$

respectively. If the upper and lower densities agree, then they are both equal to the natural density, defined as

$$\lim_{x\to\infty} \frac{|A \cap [x]|}{x}.$$

We use $X \ll Y$, $Y \gg X$, or $X = O(Y)$ to denote the estimate $|X| \leq CY$ for some absolute constant C. If there is a parameter k that goes to infinity, we use $X = o(Y)$ to denote the estimate $|X| \leq c(k)Y$ for some function $c(k)$ of that parameter that goes to zero as $k \to \infty$. Finally, we write $X \asymp Y$ for $X \ll Y \ll X$.

1.9. Lean formalization. Several of the results in this paper have also been formalized in the Lean proof assistant. The formalization was obtained with the help of the automated theorem proving tool Aristotle [Ha25], and is available at [vD]. The formalization contains fully unconditional proofs of Theorems 1.2, 1.3, 1.4, 1.5 and 1.9.

2. Basic estimates.

From the prime number theorem we have

$$\sum_{x\le p<2x}\frac{1}{p^2}\asymp\frac{1}{x\log x}$$

for all large x, and thus on summing

$$\sum_{p\ge x}\frac{1}{p^2}\asymp\frac{1}{x\log x}\ll\frac{1}{x} \tag{2.1}$$

for any $x\ge 2$. Similar arguments give the variant bounds

$$\sum_{p\ge x}\frac{1}{p(\log\log p)^2}\asymp\frac{1}{\log\log x}, \tag{2.2}$$

$$\sum_{p\ge x}\frac{\log\log p}{p^2}\asymp\frac{\log\log x}{x\log x}, \tag{2.3}$$

$$\sum_{2<p\le x}\frac{p}{(\log\log p)^2}\asymp\frac{x^2}{\log x(\log\log x)^2}, \tag{2.4}$$

$$\sum_{2<p\le x}\log\log p\asymp\frac{x\log\log x}{\log x} \tag{2.5}$$

for any $x\ge 3$.

The following basic probabilistic calculation will be used frequently:

Lemma 2.1. *Let $b \pmod W$ and $c \pmod q$ be congruence classes with W coprime to q, and let I be an interval of length $L\ge W$. Then, if n is drawn uniformly at random from those elements of $b \pmod W$ that lie in I, the probability that n lies in $c \pmod q$ is $O\big(\frac{1}{q}+\frac{W}{L}\big)$.*

Proof. Clearly, n ranges in a set of size $\asymp L/W$. Meanwhile, by the Chinese remainder theorem, the intersection of $b \pmod W$ and $c \pmod q$ is a residue class modulo Wq, which intersects I in $O\big(\frac{L}{Wq}+1\big)$ points. The claim follows. ∎

3. Property P

Proof of Theorem 1.2. We will first show part (ii). Let $f(j)$ be any function that goes to infinity with j and define $W_r := (p_1p_2\cdots p_r)^2$ with $W_0=1$, where we recall that $p_1<p_2<\cdots$ is the sequence of primes. Let $0=l_0<$

$l_1 < \cdots$ be a strictly increasing sequence of integers for which $f(j) \geq W_r$ holds for all $j \geq l_r$, which exists by the assumption that $f(j)$ goes to infinity. Finally, for a positive integer j, let $k(j)$ be such that $l_{k(j)} \leq j < l_{k(j)+1}$. Now we inductively define the sequence $A = \{1 = a_1 < a_2 < \cdots\}$ as follows. First off, $a_j = j$ for all $j \leq l_1$. Then, assuming that a_{j-1} is defined, we define a_j to be the smallest integer larger than a_{j-1} for which $a_j \equiv -r \pmod{p_r^2}$ holds for all r with $1 \leq r \leq k(j)$. We then need to show that the sequence thus defined has property P, and we need to verify that the inequality $a_j/j \leq f(j)$ holds for all $j \in \mathbb{N}$.

By the Chinese remainder theorem, we know that $a_{j'} \leq a_{j'-1} + W_{k(j')} \leq a_{j'-1} + W_{k(j)}$ holds for all $j' \leq j$, so that $a_j \leq jW_{k(j)}$ by induction. Since $f(j) \geq W_{k(j)}$, the desired inequality holds.

To see that the sequence we defined has property P, let n be any positive integer and let j be any integer larger than or equal to l_n. Then $k(j) \geq n$, so that (by using $r = n$ in the definition of a_j) we see $a_j \equiv -n \pmod{p_n^2}$, implying $a_j + n \notin \mathcal{SF}$. As n was arbitrary and j was any integer larger than or equal to l_n, we conclude that $\mathcal{SF} - n$ has finite intersection with A for all $n \in \mathbb{N}$, so that A has property P.

To prove part (i) of Theorem 1.2, assume that A is a sequence with positive upper density δ. Let C be a large absolute constant, and let H be sufficiently large depending on δ and C. We claim that there exist arbitrarily large $a \in A$ for which $a + [H]$ intersects $\mathcal{SF}$. By the pigeonhole principle, this implies that an $n \in [H]$ exists such that $n + a \in \mathcal{SF}$ for infinitely many a, establishing that A does not have property P, as desired.

It remains to establish the claim. We will use the Maier matrix method, which in this context is based on counting ([5]) the number of pairs $(a, h) \in A \times [H]$ for which $a + h \in \mathcal{SF}$. Since A has upper density δ, there exist arbitrarily large N such that

$$|A \cap [N]| \geq \delta N/2$$

and hence

$$|A \cap [\delta N/4, N]| \geq \delta N/4.$$

In particular, we can take $N \geq H$. For H large enough, we see from (1.1) and standard sieving that for any given $a \in A$, there exist $\asymp H$ choices of $h \in [H]$ such that $p^2 \nmid a + h$ for all $p \leq C$. By Lemma 2.1, for each $C < p \leq H^{1/2}$, the number of $h \in [H]$ for which $p^2 \mid a + h$ is $O(H/p^2)$. Applying (2.1), we conclude for C large enough that there are $\asymp H$ choices of $h \in [H]$ such that

([5]) One can imagine an $A \times [H]$ matrix with entry $a + h$ in the ath row and hth column. Roughly speaking, we want to sieve $a + h$ to be squarefree for many entries in a column, but initially it is easier to obtain many squarefree (or nearly squarefree) entries in a row.

$p^2 \nmid a+h$ for all $p \le H^{1/2}$. Thus

$$|\{(a,h) \in (A \cap [\delta N/4, N]) \times [H] : p^2 \nmid a+h \text{ for all } p \le H^{1/2}\}| \gg \delta NH.$$

On the other hand, for each $p > H^{1/2}$, there are $O(N/p^2)$ possible values of $a+h = O(N)$ that are divisible by p^2. Each of these is associated to at most $O(H)$ pairs (a,h), hence by (2.1) we have

$$|\{(a,h) \in (A \cap [\delta N/4, N]) \times [H] : a+h \in \mathcal{SF}\}| \gg \delta NH - O\left(\frac{NH}{H^{1/2}\log H}\right).$$

Taking C large enough, we conclude that this set is non-empty, which gives the claim. ∎

Remark 3.1. The above argument shows that the n for which $n+a \in \mathcal{SF}$ for infinitely many $a \in A$ can be chosen to be of size $O(1/(\delta^2 \log^2(1+1/\delta)))$. The bound can be improved by taking advantage of known moment estimates

$$\sum_{s_{n+1} \le x} (s_{n+1} - s_n)^\gamma \le C_\gamma x$$

for the gaps between consecutive squarefree numbers s_n, s_{n+1} and some exponent $\gamma > 1$. Any such bound can be used to bound n by $O(1/\delta^{1/(\gamma-1)})$ by applying Markov's inequality. Such a moment estimate is currently known for $\gamma < 3.75$ [Ch23], thus one can take $n \ll 1/\delta^{\frac{1}{2.75}+o(1)}$ in the limit $\delta \to 0$.

4. Property Q

4.1. The lower bound. In this section we establish Theorem 1.4. The key proposition used for this construction will be

Proposition 4.1 (Key construction). *Let $0 < \varepsilon < 1$ and let n be a sufficiently large natural number depending on ε. Then there exist arbitrarily large natural numbers n' with the following two properties:*

(i) *$n' + a \in \mathcal{SF}$ for all $a \in \mathcal{SF} \cap [n]$.*
(ii) *For any $n \le R \le n'$,*

$$\frac{|\{a \in \mathcal{SF} \cap [R] : n' + a \in \mathcal{SF}\}|}{R} \ge \frac{6}{\pi^2} - O(\varepsilon).$$

Let us assume Proposition 4.1 for the moment and conclude the proof of Theorem 1.4. By starting with a sufficiently large n_1 and applying the above proposition iteratively with $\varepsilon = 1/k$ for $k = 1, 2, \dots$, one can find a rapidly growing sequence $1 < n_1 < n_2 < \cdots$ with the following properties for all natural numbers k:

(a) $n_{k+1} + a \in \mathcal{SF}$ whenever $a \in \mathcal{SF} \cap [n_k]$.

(b) For any $n_k \leq R \leq n_{k+1}$,

$$\frac{|\{a \in \mathcal{SF} \cap [R] : n_{k+1} + a \in \mathcal{SF}\}|}{R} \geq \frac{6}{\pi^2} - O\left(\frac{1}{k}\right).$$

(c) $n_{k+1} \geq (k+1)n_k$.

We then define

$$A := \bigcup_{k=1}^{\infty} \{a \in \mathcal{SF} \cap (n_k, n_{k+1}] : n_{k+1} + a \in \mathcal{SF}\}. \tag{4.1}$$

Clearly $A \subset \mathcal{SF}$. We claim that A has property Q. Indeed, it will suffice to show that for any k, one has $n_{k+1} + a \in \mathcal{SF}$ for all $a < n_{k+1}$. For $a \leq n_k$ this follows from property (a) (since $A \subset \mathcal{SF}$) and for $n_k < a < n_{k+1}$ this follows by construction from (4.1).

Since $A \subset \mathcal{SF}$, it follows from both (1.2) and Theorem 1.3 that A has upper density at most $6/\pi^2$. To obtain the conclusion of Theorem 1.4, it thus suffices to show that A has lower density at least $6/\pi^2$. It will suffice to show that for all $k \geq 2$ and all $n_k \leq R \leq n_{k+1}$, one has the lower bound

$$\frac{|A \cap [R]|}{R} \geq \frac{6}{\pi^2} - o(1) \tag{4.2}$$

as $k \to \infty$. From (4.1) we have

$$\begin{aligned} |\{A \cap [R]\}| \geq\ & |\{a \in \mathcal{SF} \cap [R] : n_{k+1} + a \in \mathcal{SF}\}| \\ & - |\{a \in \mathcal{SF} \cap [n_k] : n_k + a \notin \mathcal{SF}\}| - n_{k-1} \end{aligned}$$

for $k \geq 2$. From property (b) we have

$$|\{a \in \mathcal{SF} \cap [R] : n_{k+1} + a \in \mathcal{SF}\}| \geq \frac{6}{\pi^2} R - o(R)$$

and

$$|\{a \in \mathcal{SF} \cap [n_k] : n_k + a \in \mathcal{SF}\}| \geq \frac{6}{\pi^2} n_k - o(n_k)$$

and hence by (1.2),

$$|\{a \in \mathcal{SF} \cap [n_k] : n_k + a \notin \mathcal{SF}\}| = o(n_k) = o(R).$$

Also by property (c) we have $n_{k-1} \leq n_k/k = o(R)$. Combining these bounds we obtain (4.2).

It remains to establish Proposition 4.1. Let ε and n be as in the proposition. Set

$$W := \prod_{p \leq n^2} p^2$$

and let x be sufficiently large depending on ε and n. Let n' be chosen uniformly at random among the multiples of W in $[x/2, x]$. It will suffice to show that n' obeys properties (i) and (ii) of the proposition with probability $1 - o(1)$ as $n \to \infty$.

We begin with (i). If $a \in \mathcal{SF}$, then for any prime $p \leq n^2$, p^2 divides W and hence n', but not $n' + a$. Also, if $a \leq n$, then $n' + a \leq 2x$. Thus, if $n' + a \notin \mathcal{SF}$, then $n' + a$ must be divisible by p^2 for some $n^2 < p \leq \sqrt{2x}$. On the other hand, from Lemma 2.1 we see that for each such p and any $a \leq n$, the probability that $n + a$ is divisible by p^2 is $O(\frac{1}{p^2} + \frac{W}{x})$. By the union bound, the probability that (i) fails is then at most

$$\ll \sum_{a=1}^{n} \sum_{n^2 < p \leq \sqrt{2x}} \left(\frac{1}{p^2} + \frac{W}{x} \right) \ll \frac{1}{n} + \frac{Wn}{\sqrt{x}} \ll \frac{1}{n}$$

thanks to (2.1) and the assumption that x is sufficiently large. Claim (i) follows.

Now we turn to (ii) and we aim to show that the inequality

$$|\{a \in \mathcal{SF} \cap [R] : n' + a \in \mathcal{SF}\}| \geq \frac{6}{\pi^2} R - O(\varepsilon R)$$

holds for all $n \leq R \leq x$ with high probability. By the error term of $O(\varepsilon R)$, we may assume without loss of generality that R is a power of $1 + \varepsilon$. And by a telescoping series, it will then suffice to show that with probability $1 - o(1)$, one has

$$|\{a \in \mathcal{SF} \cap (R, (1+\varepsilon)R] : n' + a \in \mathcal{SF}\}| \geq \frac{6}{\pi^2} \varepsilon R - O(\varepsilon^2 R)$$

for all $n \leq R \leq x$ that are powers of $1 + \varepsilon$.

So let R be a fixed power of $+\varepsilon$. From (1.2), we have

$$|\mathcal{SF} \cap (R, (1+\varepsilon)R]| = \frac{6}{\pi^2} \varepsilon R + O(\varepsilon^2 R). \tag{4.3}$$

As before, for any $a \in \mathcal{SF} \cap (R, (1+\varepsilon)R]$, if $n' + a \notin \mathcal{SF}$, then $n' + a$ must be divisible by p^2 for some $n^2 < p \leq \sqrt{2x}$. Each prime p with $n^2 < p \leq \sqrt{R}$ removes at most $O(R/p^2)$ elements from this set, for a total of at most

$$\ll \sum_{n^2 < p \leq \sqrt{R}} \frac{R}{p^2} \ll \frac{R}{n^2} \ll \varepsilon^2 R$$

elements (this claim is vacuously true if $\sqrt{R} \leq n^2$), thanks to (2.1). So it remains to consider the event that for some a in (4.3), $n' + a$ is divisible by p^2 for some $\max(n^2, \sqrt{R}) < p \leq \sqrt{2x}$. By Lemma 2.1 and the union bound, this event holds for a given R with probability

$$\ll \sum_{a=1}^{n} \sum_{\max(n^2, \sqrt{R}) < p \leq \sqrt{2x}} \left(\frac{1}{p^2} + \frac{W}{x} \right) \ll \frac{n}{\max(n^2, \sqrt{R})} + \frac{Wn}{x}$$

thanks to (2.1). Summing over $n \leq R \leq x$ that are powers of $1 + \varepsilon$, the

probability that (ii) fails is then

$$\ll \frac{1}{\varepsilon}\left(\frac{\log n}{n} + \frac{Wn\log x}{x}\right) \ll \frac{\log n}{\varepsilon n}$$

since x is large. This concludes the proof of Proposition 4.1 and hence Theorem 1.4.

4.2. Sufficiently rapidly growing sequences. Now we adapt the above arguments to prove Theorem 1.5. Let $c > 0$ be a small constant, and let C be sufficiently large depending on c. Suppose that $A = \{a_1 < a_2 < \cdots\}$ is an admissible sequence with

$$a_j \geq \exp(Cj/\log j) \tag{4.4}$$

for infinitely many j.

Let j be a sufficiently large integer for which (4.4) holds and set $x = a_j$. From (4.4) we then see that $j \leq c\log x \log\log x$. Now, with

$$W := \prod_{p \leq \frac{1}{10}\log x} p^2,$$

we have $W = x^{1/5+o(1)}$ as $x \to \infty$, by the prime number theorem. Moreover, by the Chinese remainder theorem and admissibility, there exists an integer b such that the sequence A avoids the residue class $b \pmod{p^2}$ for all $p \mid W$. Fix such a b, and let n be chosen uniformly at random from the elements of the residue class $-b \pmod W$ in $[x/2, x]$. We then aim to show that there is a positive probability that $n + a \in \mathcal{SF}$ for all $a \in A \cap [n] \subseteq A \cap [x]$.

By construction, $n + a$ is not divisible by p^2 for any $p \leq \frac{1}{10}\log x$ and $a \in A$. For $a \leq x$ we get $n + a \leq 2x$, and such numbers are also not divisible by p^2 for any $p > \sqrt{2x}$. So the only remaining primes p to check are those with $\frac{1}{10}\log x < p \leq \sqrt{2x}$. By Lemma 2.1 and the union bound, the probability that $n + a$ is not in $\mathcal{SF}$ for some $a \in A \cap [x]$, is at most

$$\ll j \sum_{\frac{1}{10}\log x < p \leq \sqrt{2x}} \left(\frac{1}{p^2} + \frac{W}{x}\right)$$

which by (2.1), $W = x^{1/5+o(1)}$ and the inequality $j \leq c\log x\log\log x$ can be bounded by

$$\ll (c\log x\log\log x)\left(\frac{1}{\log x\log\log x} + \frac{x^{7/10+o(1)}}{x}\right) \ll c.$$

For c small enough and j (and therefore x) large enough, this failure probability is smaller than 1, hence an $n \in [x/2, x]$ exists for which $n + a \in \mathcal{SF}$ for all $a \in A$ with $a < n$. Since by assumption there are infinitely many j for which (4.4) holds, we see that A has property Q as claimed.

4.3. Explicit version. In Section 1 we claimed that one can take C in Theorem 1.5 to be any constant larger than 4. To see this, we first note that we can improve (2.1) and the upper bound probability in Lemma 2.1 to $\sum_{p\geq x}\frac{1}{p^2}=\frac{1+o(1)}{x\log x}$ and $\frac{1}{q}+O(\frac{W}{L})$ respectively. Secondly, we remark that it is possible to increase the constant $\frac{1}{10}$ that appears in the definition of W to $\frac{1}{4}-\varepsilon$. With $c=\frac{1}{4}-O(\varepsilon)$ and $C=\frac{1}{c}+O(\varepsilon)$, one can then check that the final calculation to upper bound the failure probability (which currently gives $\ll c$) instead gives a quantity that is, for sufficiently large x, smaller than 1.

Furthermore, recall that in the proof of Theorem 1.5 we chose n in the interval $[x/2,x]$. If we instead use the interval $(x,x+x^{10/11})$, apply the aforementioned slightly improved versions of (2.1) and Lemma 2.1, and use the assumption that $a_j\geq\max(\exp(5j/\log j),a_{j-1}+a_{j-1}^{10/11})$ holds for all sufficiently large j, Theorem 1.6 follows analogously.

4.4. A fast growing admissible sequence that does not obey Q. We now show Theorem 1.7. With $C>0$ a large constant, we construct a random subset $A=\{a_1<a_2<\cdots\}$ of $\mathcal{SF}$ by having each squarefree $n\geq 3$ independently lie in A with probability $\mu_n:=\min\big(C\frac{\log n\log\log n}{n},1\big)$. As A is a subset of $\mathcal{SF}$, it is admissible. It will then suffice to show the following claims:

(i) Almost surely, one has $a_j\geq\exp(cj^{1/2}/\log^{1/2}j)$ for all sufficiently large j and some constant $c>0$ depending on C.
(ii) Almost surely, for sufficiently large x that are powers of two, there does not exist any $n\in[x,2x]$ such that $n+a$ is squarefree for all $a\in A\cap[x]$. In particular, A does not obey property Q.

We begin with (i). For sufficiently large x, the cardinality $|A\cap[x]|$ is a sum of independent Bernoulli variables, with mean

$$\sum_{n\in\mathcal{SF}\cap[x]}\mu_n\asymp(C+o(1))\log^2 x\log\log x$$

and variance

$$\sum_{n\in\mathcal{SF}\cap[x]}\mu_n(1-\mu_n)\asymp(C+o(1))\log^2 x\log\log x.$$

Applying Bennett's inequality ([6]) [Be62], we conclude that

$$\mathbf{P}(|A\cap[x]|\geq C^2\log^2 x\log\log x)\leq\exp(-C\log^2 x\log\log x)$$

if C is large enough. Applying the Borel–Cantelli lemma, we conclude that

([6]) One could also use Bernstein's inequality or Azuma's inequality here if desired.

almost surely, one has

$$|A \cap [x]| < C^2 \log^2 x \log\log x$$

for all sufficiently large x. In other words, if x is sufficiently large, then we almost surely have $a_j > x$ whenever $j > C^2 \log^2 x \log\log x$. Claim (i) then follows by a routine calculation.

Now we turn to (ii). Let x be a sufficiently large power of two and suppose that $A \cap [x]$ occupies every non-zero residue class modulo p^2 for each prime $p \le \log x$. Then, if n has the property that $n + a$ is squarefree for all $a \in A \cap [x]$, it is necessary for n to be congruent to 0 (mod p^2) for all $p \le \log x$. In particular, n must be divisible by $\prod_{p \le \log x} p^2$, so that no element $n \in [x, 2x]$ can have this property by the prime number theorem.

It will therefore suffice to prove that, almost surely, $A \cap [x]$ occupies every non-zero residue class modulo p^2 for each prime $p \le \log x$, for all sufficiently large powers of two x. Any given non-zero residue class b (mod p^2) avoids $A \cap [x]$ with probability

$$\begin{aligned}\prod_{\substack{n \in [x] \cap \mathcal{SF} \\ n \equiv b \,(\mathrm{mod}\ p^2)}} (1 - \mu_n) &< \prod_{\substack{n \in [\sqrt{x}, x] \cap \mathcal{SF} \\ n \equiv b \,(\mathrm{mod}\ p^2)}} \exp(-\mu_n) \\ &= \exp\biggl(- \sum_{\substack{n \in [\sqrt{x}, x] \cap \mathcal{SF} \\ n \equiv b \,(\mathrm{mod}\ p^2)}} \frac{C \log n \log\log n}{n} \biggr) \\ &\le \exp\biggl(- \sum_{\substack{n \in [\sqrt{x}, x] \cap \mathcal{SF} \\ n \equiv b \,(\mathrm{mod}\ p^2)}} \frac{C \log x \log\log x}{4n} \biggr).\end{aligned}$$

By standard sieving, for any $\sqrt{x} \le y \le x$, the number of squarefree elements in $[y, 2y]$ that are congruent to b (mod p^2) is $\gg y/p^2$. We can therefore upper bound the above probability by

$$\exp(-5 \log^2 x \log\log x / p^2) \le \log^{-5} x,$$

by setting C large enough. Since the number of non-zero residue classes b (mod p^2) is smaller than $p^2 \le \log^2 x$, while the number of primes p is smaller than $\log x$, by the union bound we conclude that $A \cap [x]$ fails to occupy every non-zero residue class modulo p^2 for each prime $p \le \log x$ with probability smaller than $\log^{-2} x$. By summing the failure probability over all sufficiently large powers of two, claim (ii) follows from the Borel–Cantelli lemma.

5. Property $\overline{P}$. We now prove Theorem 1.9, starting with part (i). If A obeys either property $\overline{P}$ or $\overline{P}_\infty$, then there exist $n_1 < n_2$ such that $n_1 + a, n_2 + a \in \mathcal{SF}$ for all but finitely many $a \in A$. Thus, after deleting

finitely many elements of A, the remaining elements avoid at least one residue class modulo p^2 for all primes p, and at least two residue classes modulo p^2 for all $p > n_2 - n_1$. From standard sieve bounds we conclude that the upper density of A is strictly less than (1.1), giving (i). In fact, by 2.1 we obtain a more precise upper bound

$$\frac{6}{\pi^2} - \frac{c}{(n_2 - n_1)\log(n_2 - n_1 + 1)}$$

for this density, for some absolute constant $c > 0$.

Now we prove (ii). We need the following lemma:

LEMMA 5.1. *For any sufficiently large $P \geq 3$, there exist arbitrarily large natural numbers n such that*

(a) $n \equiv 0 \pmod{p^2}$ *whenever* $p \leq P$,
(b) $n + a \not\equiv 0 \pmod{p^2}$ *whenever* $p > P$ *and* $1 \leq a \leq \frac{p}{(\log\log p)^2}$.

Proof. Write $W := \prod_{p \leq P} p^2$, and let x be sufficiently large depending on W. We select n uniformly at random from the elements of $[x/2, x]$ that are congruent to 0 (mod W). Clearly (a) holds. Since $n \leq x$, claim (b) is automatic for $p \geq \sqrt{2x}$ since $n + a < p^2$ for such p, for any $a \leq \frac{p}{(\log\log p)^2}$. By Lemma 2.1, the probability that p^2 divides $n + a$ for a given $P < p \leq \sqrt{2x}$ and a is $O(1/p^2 + W/x)$, thus the total failure probability of (b) is

$$\ll \sum_{P < p \leq \sqrt{2x}} \left(\frac{1}{p(\log\log p)^2} + \frac{Wp}{x(\log\log p)^2} \right),$$

which by (2.2), (2.4) can be bounded by

$$\ll \frac{1}{\log\log P} + \frac{W}{\log x(\log\log x)^2}.$$

For P and x large enough this will be smaller than 1, giving the claim. ∎

With $K \geq 3$ sufficiently large (but otherwise arbitrary) we can, by Lemma 5.1, find an infinite sequence $n_1 < n_2 < \cdots$ such that

(a) $n_j \equiv 0 \pmod{p^2}$ whenever $p \leq K \exp\exp j$,
(b) $n_j + a \not\equiv 0 \pmod{p^2}$ whenever $p > K \exp\exp j$ and $1 \leq a \leq \frac{p}{(\log\log p)^2}$.

Now set A to be the set of all natural numbers a such that $n_j + a \in \mathcal{SF}$ for all j. This set obeys properties $\overline{P}$ and $\overline{P}_\infty$ by definition, so it suffices to show that

$$|A \cap [x]| \geq \frac{6}{\pi^2} x - o(x) - O\left(\frac{x \log\log K}{K \log K} \right).$$

Indeed, as K is arbitrary one can make $O\big(\frac{x \log\log K}{K \log K}\big)$ smaller than εx by choosing K sufficiently large.

Suppose that $a \in \mathcal{SF} \cap [x]$ fails to lie in A for some $x \geq 3$. In other words, there must be a prime p and an index j such that p^2 divides $n_j + a$. By property (a) and the squarefree nature of a, this forces

$$p > K \exp\exp j,$$

so in particular $p > K \geq 3$ and

$$j < \log\log p.$$

By property (b), this then forces

$$x \geq a > \frac{p}{(\log\log p)^2}$$

and thus

$$p \ll x(\log\log x)^2.$$

We conclude that A contains all elements of $\mathcal{SF} \cap [x]$, after removing fewer than $\log\log p$ residue classes modulo p^2 for all $K < p \ll x(\log\log x)^2$. By (1.2) we get a lower bound of

$$|A \cap [x]| \geq \frac{6}{\pi^2}x - o(x) - \sum_{K<p\ll x(\log\log x)^2} O\bigg(\frac{x\log\log p}{p^2} + \log\log p\bigg),$$

which by (2.3) and (2.5) simplifies to

$$\begin{aligned} |A \cap [x]| &\geq \frac{6}{\pi^2}x - o(x) - O\bigg(\frac{x\log\log K}{K\log K} + \frac{x(\log\log x)^3}{\log x}\bigg) \\ &= \frac{6}{\pi^2}x - o(x) - O\bigg(\frac{x\log\log K}{K\log K}\bigg), \end{aligned}$$

finishing the proof.

6. Size of admissible sets. In this section we prove Theorem 1.10, beginning with the upper bound. Applying Theorem A.2 with $\omega(p^2) = 1$ for all p, we get

$$A(x) \leq \frac{x + Q^4}{\sum_{q\leq Q} h(q)}$$

for any $x, Q \geq 1$, where $h(q) = \mu^2(q)\prod_{p\mid q}\frac{1}{p^2-1}$. Since $h(q) = O(1/q^2)$ and

$$\sum_{q\geq 1} h(q) = \prod_{p\geq 2}\bigg(1 + \frac{1}{p^2-1}\bigg) = \zeta(2) = \frac{\pi^2}{6},$$

we have $\sum_{q\leq Q} h(q) = \frac{\pi^2}{6} - O(1/Q)$, and hence

$$A(x) \leq \frac{6}{\pi^2}x + O\bigg(\frac{x}{Q} + Q^4\bigg).$$

Setting $Q \coloneqq x^{1/5}$, we obtain the claim.

Now we turn to the lower bound. Let W be the product of all primes $p \leq \sqrt{x}$, and let n be chosen uniformly at random from $\mathbb{Z}/W\mathbb{Z}$. For any shift $a \in \mathbb{Z}$, the probability that $n+a$ is not divisible by any p^2 with $p \leq \sqrt{x}$ is exactly

$$\prod_{p \leq \sqrt{x}} \left(1 - \frac{1}{p^2}\right),$$

thanks to the Chinese remainder theorem. In particular, by (1.1) and (2.1), this probability is at least $\frac{6}{\pi^2} + \frac{c}{\sqrt{x}\log x}$ for some absolute constant $c > 0$. Thus, the first moment $\mathbf{E}|A|$ of the size of the random set

$$A := \{a \in [x] : p^2 \nmid n + a \text{ for all } p \leq \sqrt{x}\}$$

is at least $\frac{6}{\pi^2}x + \frac{c\sqrt{x}}{\log x}$. Thus we can find an n for which

$$|A| \geq \frac{6}{\pi^2}x + \frac{c\sqrt{x}}{\log x}.$$

By construction, this set avoids a residue class modulo p^2 for all $p \leq \sqrt{x}$. On the other hand, for all $p > \sqrt{x}$ we see that $A \subseteq [x]$ avoids the residue class $\lfloor x \rfloor + 1 \pmod{p^2}$. Hence A is admissible, giving the lower bound

$$A(x) \geq \frac{6}{\pi^2}x + \frac{c\sqrt{x}}{\log x}$$

as required.

Remark 6.1. It seems likely that by calculating higher moments of $|A|$ one could obtain some slight improvements to the lower bound (thereby potentially determining whether $A(x) > |\mathcal{SF} \cap [x]|$ holds for all $x \geq 18$), but we will not perform these calculations here.

Appendix A. A large sieve inequality for square moduli. We recall the large sieve inequality of Montgomery and Vaughan [MV73]:

Theorem A.1. *Let M and $N \geq 1$ be integers, and let A be a subset of the interval $[M+1, M+N]$ that avoids $\omega(p) < p$ residue classes modulo p for each prime p. Then for any $Q \geq 1$, one has*

$$|A| \leq \frac{N + Q^2}{\sum_{q \leq Q} h(q)}, \quad \text{where} \quad h(q) := \mu^2(q) \prod_{p \mid q} \frac{\omega(p)}{p - \omega(p)}.$$

Indeed, this is [MV73, Corollary 1] after replacing [MV73, (1.6)] in the proof of that corollary with [MV73, (1.4)]. In this appendix we prove the following variant to Theorem A.1:

THEOREM A.2. *Let M and $N \geq 1$ be integers, and let A be a subset of the interval $[M+1, M+N]$ that avoids $\omega(p^2) < p^2$ residue classes modulo p^2 for each prime p. Then for any $Q \geq 1$, one has*

$$|A| \leq \frac{N+Q^4}{\sum_{q\leq Q} h(q)}, \quad \text{where} \quad h(q) := \mu^2(q) \prod_{p|q} \frac{\omega(p^2)}{p^2-\omega(p^2)}.$$

This result was essentially established in [Ga74] (with a slightly different bound) under the additional assumption that the $\omega(p^2)$ residue classes removed had a tree structure, coming from first removing some residue classes modulo p, and then removing the same number of residue classes modulo p^2 for each surviving residue class modulo p. However, we cannot assume this tree structure in our applications, so we need Theorem A.2 as a substitute. One could also generalize this result to higher powers than squares, with the obvious modifications to the statement of the theorem; we leave this to the interested reader.

Proof of Theorem A.2. Let a_n be any sequence of complex numbers supported on A, and form the exponential sums $S(\alpha) = \sum_{n\in A} a_n e(n\alpha)$ with $e(\alpha) := e^{2\pi i\alpha}$. We then claim the inequality

$$\sum_{\substack{q|q'\\ q'|q^2}} \sum_{\substack{1\leq a\leq q'\\ (a,q')=1}} \left|S\left(\frac{a}{q'}+\alpha\right)\right|^2 \geq h(q)|S(\alpha)|^2 \tag{A.1}$$

for any squarefree q and any $\alpha \in \mathbb{R}$. By the usual Chinese remainder theorem argument (see, e.g., [Ga74, p. 493]), if this inequality is proven for $q = q_1$ and $q = q_2$ for some coprime q_1, q_2, then it also holds for $q = q_1 q_2$. As one can check (A.1) by hand if $q = 1$, it therefore suffices to verify the inequality for $q = p$ a prime, in which case the outer sum only contains the two terms corresponding to $q' = p$ and $q' = p^2$. By modulating the sequence a_n by $e(n\alpha)$ we may furthermore normalize $\alpha = 0$, so that we get

$$\begin{aligned}\sum_{\substack{q|q'\\ q'|q^2}} \sum_{\substack{1\leq a\leq q'\\ (a,q')=1}} \left|S\left(\frac{a}{q'}+\alpha\right)\right|^2 &= \sum_{1\leq a<p} \left|S\left(\frac{a}{p}\right)\right|^2 + \sum_{\substack{1\leq a<p^2\\ a\nmid p}} \left|S\left(\frac{a}{p^2}\right)\right|^2 \\ &= \sum_{1\leq a<p^2} \left|S\left(\frac{a}{p^2}\right)\right|^2.\end{aligned}$$

Since $h(p) = \frac{\omega(p^2)}{p^2-\omega(p^2)}$, in order to deduce A.1 it remains to prove

$$\sum_{1\leq a<p^2} \left|S\left(\frac{a}{p^2}\right)\right|^2 \geq \frac{\omega(p^2)}{p^2-\omega(p^2)}|S(0)|^2. \tag{A.2}$$

Now, if we let $\nu(n) := \omega(p^2) - p^2 1_S$, where S is the union of the $\omega(p^2)$ residue classes modulo p^2 one is removing, then $\nu(n)$ is p^2-periodic and of

mean zero, hence admits a Fourier expansion

$$\nu(n) = \sum_{1 \le a < p^2} c_a e(an/p^2)$$

for some coefficients c_a. By the Plancherel identity and a brief calculation one has

$$\sum_{1 \le a < p^2} |c_a|^2 = (p^2 - \omega(p^2))\omega(p^2),$$

and so by Cauchy–Schwarz we have

$$\left| \sum_{1 \le a < p^2} c_a S\left(\frac{a}{p^2}\right) \right|^2 \le (p^2 - \omega(p^2))\omega(p^2) \sum_{1 \le a < p^2} \left| S\left(\frac{a}{p^2}\right) \right|^2.$$

However, the left-hand side can be rewritten as

$$\left| \sum_{n \in A} \nu(n) a_n \right|^2 = \omega(p^2)^2 |S(0)|^2$$

so that A.2 and hence A.1 follow.

Summing (A.1) with $\alpha = 0$ and squarefree $q \le Q$, and $a_n = 1_A(n)$, we conclude that

$$\sum_{q \le Q} \mu^2(q) \sum_{\substack{q|q' \\ q'|q^2}} \sum_{\substack{1 \le a \le q' \\ (a,q')=1}} \left| S\left(\frac{a}{q'}\right) \right|^2 \ge |A| \sum_{q \le Q} h(q).$$

The fractions $\frac{a}{q'}$ in the sum on the left-hand side are $\frac{1}{Q^4}$-separated, and so by the arithmetic large sieve inequality [MV73, Theorem 1] one has

$$\sum_{q \le Q} \mu^2(q) \sum_{\substack{q|q' \\ q'|q^2}} \sum_{\substack{1 \le a \le q' \\ (a,q')=1}} \left| S\left(\frac{a}{q'}\right) \right|^2 \le N + Q^4,$$

finishing the proof. ∎

Acknowledgements. We thank Thomas Bloom for his website [Bl] where the above problems were brought to the authors' attention at [Bl, Problems 1102, 1103]. We also thank Benjamin Bedert for references concerning sets with squarefree sums, and we thank the anonymous referee for helpful suggestions.

Funding. TT was supported by the James and Carol Collins Chair, the Mathematical Analysis & Application Research Fund, and by NSF grant DMS-2347850, and is particularly grateful to recent donors to the Research Fund.

References

[Be62] G. Bennett, *Probability inequalities for the sum of independent random variables*, J. Amer. Statist. Assoc. 57 (1962), 33–45.

[Bl] T. F. Bloom, *Erdős problems*, https://www.erdosproblems.com (accessed November 3, 2025).

[Ch23] T. H. Chan, *On moments of gaps between consecutive square-free numbers*, Moscow J. Combin. Number Theory 12 (2023), 287–295.

[vD] W. van Doorn, *ErdosProblem1102.lean*, GitHub repository, https://github.com/Woett/Lean-files/blob/main/ErdosProblem1102.lean.

[Er81] P. Erdős, *Some problems and results on additive and multiplicative number theory*, in: Analytic Number Theory (Philadelphia, PA, 1980), Lecture Notes in Math. 899, Springer, Berlin, 1981, 171–182.

[ES87] P. Erdős and A. Sárközy, *On divisibility properties of integers of the form $a + a'$*, Acta Math. Hungar. 50 (1987), 117–122.

[Fi93] M. Filaseta, *On the distribution of gaps between squarefree numbers*, Mathematika 40 (1993), 88–101.

[Ga74] P. X. Gallagher, *Sieving by prime powers*, Acta Arith. 24 (1974), 491–497.

[Gy01] K. Gyarmati, *On divisibility properties of integers of the form $ab + 1$*, Period. Math. Hungar. 43 (2001), 71–79.

[Ha25] The Harmonic Team, *Aristotle: IMO-level automated theorem proving*, arXiv:2510.01346v1 (2025).

[Ko04] S. V. Konyagin, *Problems of the set of square-free numbers*, Izv. Ross. Akad. Nauk Ser. Mat. 68 (2004), no. 3, 63–90 (in Russian); English transl.: Izv. Math. 68 (2004), no. 3, 493–520.

[LNS90] C. B. Lacampagne, C. A. Nicol and J. L. Selfridge, *Sets with nonsquarefree sums*, in: Number Theory (Banff, AB, 1988), de Gruyter, Berlin, 1990, 299–311.

[Li16] H.-Q. Liu, *On the distribution of squarefree numbers*, J. Number Theory 159 (2016), 202–222.

[MV73] H. L. Montgomery and R. C. Vaughan, *The large sieve*, Mathematika 20 (1973), 119–134.

[Na89] M. B. Nathanson, *Sumsets containing k-free integers*, in: Number Theory (Ulm, 1987), Lecture Notes in Math. 1380, Springer, New York, 1989, 179–184.

[Sa92] G. N. Sárközy, *On a problem of P. Erdős*, Acta Math. Hungar. 60 (1992), 271–282.

[Sc05] T. Schoen, *Squarefree numbers in sumsets*, Indag. Math. (N.S.) 16 (2005), 251–265.

[TZ23] T. Tao and T. Ziegler, *Infinite partial sumsets in the primes*, J. Anal. Math. 151 (2023), 375–389.

[Wa63] A. Walfisz, *Weylsche Exponentialsummen in der neueren Zahlentheorie*, Math. Forschungsber. 16, Deutscher Verlag der Wiss., Berlin, 1963.

Wouter van Doorn
Groningen, The Netherlands
E-mail: wonterman1@hotmail.com

Terence Tao
Department of Mathematics
University of California, Los Angeles
Los Angeles, CA, USA
E-mail: tao@math.ucla.edu