On prefix–free and suffix–free sequences of integers

Rudolf Ahlswede and Levon H. Khachatrian
Universität Bielefeld, Fakultät für Mathematik,
Postfach 100131, D–33501 Bielefeld, Germany

and

András Sárközy*
Eötvös University, Department of Algebra and Number Theory,
H-1088 Budapest, Muzeum krt. 6–8,Hungary

* Research partially supported by the Hungarian National Foundation for Scientific Research, Grant no. T017433. This paper was written while the third author was visiting the Universität Bielefeld.

1 Introduction

The set of the positive integers and positive square–free integers are denoted by $\mathbb{N}$ and $\mathbb{N}^*$, respectively, and we write $\mathbb{N}(n) = \mathbb{N} \cap [1, n], \mathbb{N}^*(n) = \mathbb{N}^* \cap [1, n]$, where $[1, n] = \{1, 2, \ldots, n\}$. The set of primes is denoted by $\mathcal{P}$. The smallest and greatest prime factors of the positive integer n are denoted by $p(n)$ and $P(n)$, respectively. $\omega(n)$ denotes the number of distinct prime factors of n, while $\Omega(n)$ denotes the number of prime factors of n counted with multiplicity:

$$\omega(n) = \sum_{p|n} 1, \quad \Omega(n) = \sum_{p^\alpha \| n} \alpha.$$

$\mu(n)$ denotes the Möbius function.

The counting function of a set $\mathcal{A} \subset \mathbb{N}$, denoted by $A(x)$, is defined by

$$A(x) = |\mathcal{A} \cap [1, x]|.$$

The upper density $\overline{d}(\mathcal{A})$ and the lower density $\underline{d}(\mathcal{A})$ of the infinite set $\mathcal{A} \subset \mathbb{N}$ are defined by

$$\overline{d}(\mathcal{A}) = \limsup_{x\to\infty} \frac{A(x)}{x}$$

and

$$\underline{d}(\mathcal{A}) = \liminf_{x\to\infty} \frac{A(x)}{x},$$

respectively, and if $\overline{d}(\mathcal{A}) = \underline{d}(\mathcal{A})$, then the density $d(\mathcal{A})$ of $\mathcal{A}$ is defined as

$$d(\mathcal{A}) = \overline{d}(\mathcal{A}) = \underline{d}(\mathcal{A}).$$

The upper logarithmic density $\overline{\delta}(\mathcal{A})$ of the infinite set $\mathcal{A} \subset \mathbb{N}$ is defined by

$$\overline{\delta}(\mathcal{A}) = \limsup_{x\to\infty} \frac{1}{\log x} \sum_{\substack{a\in\mathcal{A} \\ a\le x}} \frac{1}{a},$$

and the definitions of the lower logarithmic density $\underline{\delta}(\mathcal{A})$ and logarithmic density $\delta(\mathcal{A})$ are similar.

For $\mathcal{A} \subset \mathbb{N}$, $s > 1$ write

$$f_{\mathcal{A}}(s) = \sum_{a\in\mathcal{A}} a^{-s}.$$

Then the lower and upper Dirichlet densities of $\mathcal{A}$ are defined by

$$\underline{D}(\mathcal{A}) = \liminf_{s\to 1^+} (s-1) f_{\mathcal{A}}(s)$$

and

$$\overline{D}(\mathcal{A}) = \limsup_{s\to 1^+} (s-1) f_{\mathcal{A}}(s),$$

respectively. If $\overline{D}(\mathcal{A}) = \underline{D}(\mathcal{A})$, then the Dirichlet density $D(\mathcal{A})$ of $\mathcal{A}$ is defined as

$$D(\mathcal{A}) = \overline{D}(\mathcal{A}) = \underline{D}(\mathcal{A}).$$

It is known that for every $\mathcal{A} \subset \mathbb{N}$ we have

$$\overline{\delta}(\mathcal{A}) = \overline{D}(\mathcal{A}), \underline{\delta}(\mathcal{A}) = \underline{D}(\mathcal{A})$$

and

$$0 \le \underline{d}(\mathcal{A}) \le \underline{\delta}(\mathcal{A}) \le \overline{\delta}(\mathcal{A}) \le \overline{d}(\mathcal{A}) \le 1.$$

We will study mostly sets of square–free integers. It is well–known that

$$d(\mathbb{N}^*) = \frac{6}{\pi^2}. \tag{1.1}$$

We will compare the density of a set $\mathcal{A} \subset \mathbb{N}^*$ with the density of $\mathbb{N}^*$, and the density obtained in this way will be denoted by an asterisque. Thus, e.g., for $\mathcal{A} \subset \mathbb{N}^*$ we write

$$d^*(\mathcal{A}) = \frac{d(\mathcal{A})}{d(\mathbb{N}^*)} = \frac{\pi^2}{6} d(\mathcal{A}),$$

$$\underline{\delta}^*(\mathcal{A}) = \frac{\underline{\delta}(\mathcal{A})}{\delta(\mathbb{N}^*)} = \frac{\pi^2}{6} \underline{\delta}(\mathcal{A}),$$

etc.

A set $\mathcal{A} \subset \mathbb{N}$ is said to be *primitive* if there are no a, a' with $a \in \mathcal{A}$, $a' \in \mathcal{A}$, $a \ne a'$ and $a|a'$. Let $F(n)$ denote the cardinality of the greatest primitive set selected from $\{1, 2, \dots, n\}$. Then it is easy to see [9] that

$$F(n) = \left\lceil \frac{n}{2} \right\rceil \left(= \left(\frac{1}{2} + o(1) \right) n \right). \tag{1.2}$$

By the results of Besicovitch [3] and Erdös [6], for all $\varepsilon > 0$

$$\text{there is an infinite primitive set } \; \mathcal{A} \subset \mathbb{N} \; \text{ with } \; \overline{d}(\mathcal{A}) > \frac{1}{2} - \varepsilon. \tag{1.3}$$

Behrend [4] proved that if $\mathcal{A} \subset \{1, 2, \dots, N\}$ and $\mathcal{A}$ is primitive then we have

$$\sum_{a \in \mathcal{A}} \frac{1}{a} < c_1 \frac{\log N}{(\log \log N)^{1/2}} \tag{1.4}$$

(so that an infinite primitive set must have zero logarithmic density) and Erdös [5] proved that if $\mathcal{A} \subset \mathbb{N}$ is a (finite or infinite) primitive set then

$$\sum_{a \in \mathcal{A}} \frac{1}{a \log a} < c_2. \tag{1.5}$$

These results have been extended in various directions; surveys of this field are given in [2], [8], [9], [10].

Next we will introduce two notions of information theoretical background. If a, b are positive square–free integers with the property that $a|b$ and $p(b/a) > P(a)$, i.e., they are of the form $a = p_1 \dots p_r$, $b = p_1 \dots p_r p_{r+1} \dots p_t$ where $p_1 < \dots < p_r < p_{r+1} < \dots < p_t$ are distinct primes (with $t > r$), then we say that a is *prefix* of b and we write $a|_p b$. If $\mathcal{A} \subset \mathbb{N}^*$ is a set such that there are no $a \in \mathcal{A}$, $b \in \mathcal{A}$ with $a|_p b$, then $\mathcal{A}$ is said to be *prefix–free*. Similarly, if $a|b$ and $P(b/a) < p(a)$, then a is called *suffix* of b and we write $a|_s b$. If $\mathcal{A} \subset \mathbb{N}^*$ is a set such that there are no $a \in \mathcal{A}$, $b \in \mathcal{A}$ with $a|_s b$, then $\mathcal{A}$ is said to be *suffix–free*. (Both notions, prefix and suffix, could be extended to the non–squarefree case as well, however, to simplify the discussion here we restrict ourselves to the square–free case.)

A further motivation for introducing and studying these concepts is that there is a close connection between prefix–freeness and primitivity: clearly,

$$\text{if a set } \ \mathcal{A} \subset \mathbb{N} \ \text{ is primitive, then it is prefix–free.} \tag{1.6}$$

Since prefix–freeness appears in connection with primitivity (see the proof of Theorem 3 below), one might like to study how close these concepts are.

Based on these considerations, in this paper our goal is to study density related properties of prefix–free and suffix–free sets.

2 The problems and results

Our first goal is to study the "prefix–free analog" of (1.2). Let $G(n)$ denote the cardinality of the greatest prefix–free set selected from $\mathbb{N}^*(n)$, and let $P^+(a)$ denote the smallest prime greater than $P(a)$.

Theorem 1. *Write*

$$\mathcal{B}(n) = \big\{ b : b \in \mathbb{N}^*(n), bP^+(b) > n \big\}. \tag{2.1}$$

Then $\mathcal{B}(n)$ is prefix–free and

$$G(n) = |\mathcal{B}(n)|.$$

Note that it follows from the prime number theorem that, if $1 > \varepsilon > 0$ and $n > n_1(\varepsilon)$, then for all $b \in \mathbb{N}^*(n)$, $b > (1+\varepsilon)\frac{n}{\log n}$ we have

$$P(b) > \left(1 - \frac{\varepsilon}{2}\right) \log n$$

so that

$$bP^+(b) > bP(b) > (1+\varepsilon)\left(1 - \frac{\varepsilon}{2}\right) \log n > \log n$$

and thus $b \in \mathcal{B}(n)$. It follows that

$$G(n) > \left(1 - \frac{1+\varepsilon}{\log n}\right) N^*(n)$$

so that

$$\lim_{n\to\infty} \frac{G(n)}{N^*(n)} = 1; \tag{2.2}$$

compare this with (1.2).

A combination of (2.2) with result of Erdös [6] gives

Corollary 1. *For all $\varepsilon > 0$ there is an infinite prefix–free set $\mathcal{A} \subset \mathbb{N}^*$ with*

$$\overline{d}^*(\mathcal{A}) > 1 - \varepsilon.$$

Since this can be derived trivially from (2.2) by using ideas of [6], we will not present the details here.

The "prefix–free analog" of Behrend's theorem (1.4) reflects an interesting difference between primitive sets and prefix–free sets. Indeed, consider now instead of $G(n)$

$$E(n) = \max_{\text{prefix–free } \mathcal{A}\subset\mathbb{N}^*(n)} \sum_{a\in\mathcal{A}} \frac{1}{a}. \tag{2.3}$$

Theorem 2. *For every $\varepsilon > 0$ and $n > n_2(\varepsilon)$, suitable,*

$$0,2689 - \varepsilon < \frac{E(n)}{\sum\limits_{b\in\mathbb{N}^*(n)} \frac{1}{b}} < 0,7311 + \varepsilon.$$

Actually, we know for every $n \in \mathbb{N}$ the unique optimal prefix–free $\mathcal{A} \subset \mathbb{N}^*(n)$ for which $E(n)$ in (2.3) is assumed, but the value, and particularly also $\lim\limits_{n\to\infty} E(n)$, which we conjecture to exist, is hard to estimate.

We shall show that the proofs of both, Theorem 1 and Theorem 2, can be given by the same approach via the Basic Lemma 1 in Section 3 involving multiplicative functions. Actually, this lemma seems to be useful also for other cases.

For instance it shades a new light on a well–known conjecture of Erdös concerning (finite or infinite) primitive sets, which says that for every primitive set $\mathcal{A} \subset \mathbb{N}$

$$\sum_{a\in\mathcal{A}} \frac{1}{a\log a} \le \sum_{p\in\mathcal{P}} \frac{1}{p\log p}.$$

Consider now for any positive, multiplicative function f

$$L_f(\infty) = \max_{\text{prefix–free } \mathcal{A}\subset\mathbb{N}^*} \sum_{a\in\mathcal{A}} f(a) \tag{2.4}$$

then we have the

Proposition 1. Let f be a multiplicative function such that

$$\sum_{p\geq 3, p\in\mathcal{P}} f(p) < 1,$$

then $L_f(\infty)$ is assumed at the set of primes. In particular, if $f(m) = m^\alpha$, then for every $\alpha \leq \alpha_0$, where $\alpha_0 \in \mathbb{R}$ and $\sum\limits_{p\geq 3} p^{\alpha_0} = 1$, the primes are the optimal set.

Next we will extend Erdös's theorem (1.5) to prefix–free sets:

Theorem 3. *There is an absolute constant c_3 such that if $\mathcal{A} \subset \mathbb{N}^*$ is a (finite or infinite) prefix–free set, then*

$$\sum_{a\in\mathcal{A}} \frac{1}{a\log a} < c_3.$$

Indeed, in Erdös's proof [5] only the prefix property of primitive sequences is used (that they possess by (1.6)) so that it also gives the more general result Theorem 3. To see that indeed it is so, for the sake of completeness we will sketch the proof in Section 5 (leaving some technical details to the reader).

It follows easily from Theorem 3 (proving by contradiction and using partial summation) that

Corollary 2. *If $\mathcal{A} \subset \mathbb{N}^*$ is an infinite prefix–free set then we have*

$$A(x) < \frac{x}{\log\log x \ \log\log\log x} \tag{2.5}$$

for infinitely many x (and, by (1.5), if $\mathcal{A} \subset \mathbb{N}$ is primitive then (2.5) also holds infinitely often).

One might like to know how far the upper bound in (2.5) is from the best possible. This is closely related to one of the favourite problems of Erdös. In [8] this problem is formulated in the following way (and Erdös mentioned it in numerous problem papers as well):

"The following problem seems difficult: Let $b_1 < b_2 \dots$ be an infinite sequence of integers. What is a necessary and sufficient condition that there should exist a primitive sequence $a_1 < a_2 \dots$ satisfying $a_n < cb_n$ for every n?

From (1.5) ... we obtain that we must have

$$\sum_{i=1}^{\infty} \frac{1}{b_i \log b_i} < \infty \ \dots \tag{2.6}$$

We know that (2.6) is not sufficient — it is not clear whether a simple necessary and sufficient condition exists."

This is followed by a lengthy discussion of the problem how large one can make $\sum_{a \le x} \frac{1}{a}$ uniformly in x for a primitive set $a_1 < \dots$ (see also [7]).

It seems to be a more natural (although more difficult) problem to replace here the sum $\sum_{a \le x} \frac{1}{a}$ by the counting function $A(x)$, i.e. to study the problem how large one can make $A(x)$ uniformly in x for a primitive set $\mathcal{A}$ — and this is the question asked by us also for prefix–free sets. In [1] we gave a quite satisfactory answer by proving that (2.5) is best possible apart from a factor $(\log\log\log x)^{\varepsilon}$:

Theorem 4. [1] *For all $\varepsilon > 0$ there is an infinite primitive (and therefore also prefix–free) set $\mathcal{A} \subset \mathbb{N}$ such that for $a > x_0(\varepsilon)$ we have*

$$A(x) > \frac{x}{\log\log x (\log\log\log x)^{1+\varepsilon}}.$$

By a standard argument it can be shown that here $\mathcal{A} \subset \mathbb{N}$ can be replaced by $\mathcal{A}^* \subset \mathbb{N}$ (and the same lower bound holds), and by (1.6), this $\mathcal{A}^*$ also is prefix–free. Thus the behaviour of primitive and prefix–free sets is similar as far as the maximal rate of growth of the counting function is concerned: in both cases the estimates (2.5) and the one in Theorem 4 can be given.

Problem 1. Is it true that if $\mathcal{A} \subset \mathbb{N}^*$ is an infinite set with

$$\overline{\delta}^*(\mathcal{A}) > 0, \tag{2.7}$$

then $\mathcal{A}$ contains an infinite "prefix chain", i.e., there is an infinite subset $\{a_{i_1}, a_{i_2}, \dots\}$ of $\mathcal{A}$ with $a_{i_1}|_p a_{i_2}|_p a_{i_3} \dots$?

Note that by a theorem of Davenport and Erdös [11], (2.7) implies that $\mathcal{A}$ contains an infinite divisibility chain $a_{i_1}|a_{i_2}|a_{i_3}\dots$.

The finite analog of Problem 1 is easier. Indeed, we will prove in Section 6

Theorem 5.

(i) *If $n > n_3$,*

$$\mathcal{A} = \{a_1, \dots, a_t\} \subset \mathbb{N}^*(n) \tag{2.8}$$

and

$$E(\mathcal{A}) \overset{def}{=} \sum_{a \in \mathcal{A}} \frac{1}{a \log a} > c_3 \tag{2.9}$$

(where c_3 is the constant defined in Theorem 3), then, writing

$$k = \left[\frac{E(\mathcal{A})}{c_3}\right] + 1, \tag{2.10}$$

$\mathcal{A}$ contains a prefix chain of length k, i.e., there is a subset $\{a_{i_1}, a_{i_2}, \dots, a_{i_k}\}$ of $\mathcal{A}$ with $a_{i_1}|_p a_{i_2}|_p \dots a_{i_k}$.

(ii) *There are numbers c_4 and n_4 with the following properties: there is an infinite set $\mathcal{A} \subset \mathbb{N}^*$ such that*

$$d^*(\mathcal{A}) = 1 \tag{2.11}$$

and, writing

$$E(\mathcal{A}, n) = \sum_{a \in \mathcal{A}, a \le n} \frac{1}{a \log a},$$

for $n > n_4$ the set $\mathcal{A} \cap \mathbb{N}^*(n)$ does not contain a prefix chain longer than $c_4 E(\mathcal{A}, n)$.

(So that (i) is best possible apart from a constant factor in the length of the maximal chain.)

While the behaviour of prefix–free and primitive sets is similar as far as the maximal rate of growth of the counting function is concerned, the behaviour of the suffix–free sets is very much different and, indeed, they can be much "denser".

We consider now the cardinality and the asymptotic density of suffix–free sets.

Let $H(n)$ denote the cardinality of the largest suffix–free set selected from $\mathbb{N}^*(n)$.

Theorem 6. *The set*

$$\mathcal{C}(n) = \left\{ c \in \mathbb{N}^*(n) : 2|c \right\} \cup \left\{ \mathbb{N}^*(n) \cap \left(\frac{n}{2}, n \right] \right\}$$

is suffix–free and $|\mathcal{C}(n)| = H(n)$.

Corollary 3.

$$\lim_{n \to \infty} \frac{H(n)}{|\mathbb{N}^*(n)|} = \frac{2}{3}.$$

Using ideas of Besicovitch [3] and Erdös [5, 6] one can easily get the following result, whose proof is not presented in this paper.

Corollary 4. *For every $\varepsilon > 0$ there exists an infinite suffix–free set $\mathcal{C}$ such that*

$$\overline{d^*}\mathcal{C} > \frac{2}{3} - \varepsilon.$$

Finally we discuss logarithmic densities of sufix–free sets. Let

$$K(n) = \max_{\text{suffix–free } \mathcal{A} \in \mathbb{N}^*} \sum_{a \in \mathcal{A}} \frac{1}{a}.$$

In contrast to the case of prefix–free sets, here Basic Lemma 2 of Section 3 gives a very simple description of the optimal set.

Theorem 7. *Let $\mathcal{B}$ be the set from Basic Lemma 2. We have*
$\mathcal{B} = \mathcal{B}_1 \dot{\cup} \mathcal{B}_2$, *where*
$\mathcal{B}_1 = \left\{2 \cdot a, 3 \cdot a, 5 \cdot a : a \in \mathbb{N}^* \left(\frac{n}{5}\right) \text{ and } (a, 30) = 1\right\}$ *and*
$\mathcal{B}_2 = \left\{a \in \mathbb{N}^* : \frac{n}{5} < a \leq n \text{ and } (a, 30) = 1\right\}$.

Simple calculations yield

Corollary 5.

$$\lim_{n \to \infty} \frac{K(n)}{\sum\limits_{a \in \mathbb{N}^*(n)} \frac{1}{a}} = \frac{31}{72}.$$

Corollary 6.

(i) *For any infinite suffix–free set $\mathcal{C}$ holds*

$$\overline{D^*}\mathcal{C} = \overline{\delta^*}\mathcal{C} \leq \frac{31}{72}.$$

(ii) *Define*

$$\mathcal{C} = \left\{2 \cdot a, 3 \cdot a, 5 \cdot a : a \in \mathbb{N}^* \text{ and } (a, 30) = 1\right\}.$$

Then $\mathcal{C}$ is an infinite suffix–free set and

$$d^*\mathcal{C} = \frac{31}{72}.$$

Similarly to $L_f(\infty)$ for infinite prefix–free sets define the quantity $S_f(\infty)$ for infinite suffix–free sets, where f is a positive multiplicative function:

$$S_f(\infty) = \max_{\text{suffix–free } \mathcal{A} \subset \mathbb{N}^*} \sum_{a \in \mathcal{A}} f(a).$$

Proposition 2. *Let f be a multiplicative function such that $\sum_{p \in \mathcal{P}} f(p) < 1$.*

Then $S_f(\infty)$ is assumed at the set of primes. In particular, if $f(m) = m^\beta$, then for every $\beta \leq \beta_0$, where $\beta_0 \in \mathbb{R}$ and $\sum\limits_{p \in P} p^{\beta_0} = 1$, the primes are the optimal set.

Remark: We note the difference to Proposition 1, where the summation starts from $p \geq 3$, and hence clearly $\beta_0 < \alpha_0$.

3 Two basic lemmas

For any positive, multiplicative function f define

$$L_f(n) = \max_{\text{prefix-free } \mathcal{A} \in \mathbb{N}^*(n)} \sum_{a \in \mathcal{A}} f(a).$$

Basic Lemma 1. *Write*

$$\mathcal{A} = \left\{ \begin{array}{ll} a \in \mathbb{N}^*(n): & \text{(i)} \sum\limits_{P(a) < p \leq \frac{n}{a}} f(p) < 1 \text{ and} \\ & \text{(ii)} \sum\limits_{P(a') < p \leq \frac{n}{a'}} f(p) \geq 1, \text{ where } a' = \frac{a}{P(a)}, \end{array} \right\}$$

We assume that (i) *always holds if* $P(a) \geq \frac{n}{a}$ *or* $P(a) < \frac{n}{a}$*, but there is no prime in the interval* $\left(P(a), \frac{n}{a}\right]$*. We also assume that* (ii) *always holds if* $a \in \mathcal{P}$*.*

Then $\mathcal{A}$ *is prefix–free and*

$$\sum_{a \in \mathcal{A}} f(a) = L_f(n).$$

Proof: We show that $\mathcal{A}$ is prefix–free. Assume to the opposite that there are $a, b \in \mathcal{A}$ such that $a|_p b$, that is $b = a \cdot c$, $p(c) > P(a)$.

We have from condition (i) for $a \in \mathcal{A}$

$$\sum_{P(a) < p \leq \frac{n}{a}} f(p) < 1 \tag{3.1}$$

and from condition (ii) for $b' = \frac{b}{P(b)} \geq a$

$$\sum_{P(b') < p \leq \frac{n}{b'}} f(p) \geq 1. \tag{3.2}$$

Since $P(b') \geq P(a)$, $b' \geq a$ and consequently $\frac{n}{a} \geq \frac{n}{b'}$, (3.1) and (3.2) are not compatible. Hence $\mathcal{A}$ is prefix–free.

Now we show that

$$\sum_{a \in \mathcal{A}} f(a) = L_f(n). \tag{3.3}$$

Let

$$\mathcal{L}_f(n) = \left\{ \mathcal{B} \subset \mathbb{N}^*(n) : \mathcal{B} \text{ is prefix–free and } \sum_{b \in \mathcal{B}} f(b) = L_f(n) \right\}.$$

So, equivalent to (3.3) is $\mathcal{A} \in \mathcal{L}_f(n)$.

Let $\mathcal{B} \in \mathcal{L}_f(n)$ be a set for which

$$\sum_{b \in \mathcal{B}} b \quad \text{is maximal among elements of } \mathcal{L}_f(n). \tag{3.4}$$

We claim that $\mathcal{B} = \mathcal{A}$. For this we have to prove that (i) and (ii) hold for every element $b \in \mathcal{B}$. We show that (i) holds. Assume to the opposite that for an element $b \in \mathcal{B}$ we have

$$\sum_{P(b) < p \leq \frac{n}{b}} f(p) \geq 1. \tag{3.5}$$

Define

$$\mathcal{B}' = \big(\mathcal{B} \smallsetminus \{b\}\big) \cup \left\{b \cdot p : p > P(b), p \leq \frac{n}{b}\right\}.$$

Since $\mathcal{B}$ is prefix–free necessarily $b \cdot p \notin \mathcal{B}$ for all $p(b) < p \leq \frac{n}{b}$. Clearly $\mathcal{B}' \subset \mathbb{N}^*(n)$.

It is easy to see that $\mathcal{B}'$ is prefix–free and

$$\sum_{b \in \mathcal{B}} b < \sum_{b \in \mathcal{B}'} b. \tag{3.6}$$

Moreover, since f is a multiplicative function, we have

$$\sum_{P(b) < p \leq \frac{n}{b}} f(p \cdot b) = f(b) \cdot \sum_{P(b) < p \leq \frac{n}{b}} f(p) \geq f(b) \text{ (by assymption (3.5))}$$

and consequently

$$\sum_{b \in \mathcal{B}} f(b) \leq \sum_{b \in \mathcal{B}'} f(b). \tag{3.7}$$

Hence $\mathcal{B}' \in \mathcal{L}_f(n)$, which is a contradiction (see (3.4) and (3.6)).

Therefore for all $b \in \mathcal{B}$ (i) holds.

Now we show that for all $b \in \mathcal{B}$ (ii) holds. Assume to the opposite that for a $b \in \mathcal{B}$ we have

$$\sum_{P(b') < p \leq \frac{n}{b'}} f(p) < 1, \text{ where } b' = \frac{b}{P(b)}. \tag{3.8}$$

Among such elements $b \in \mathcal{B}$ we choose one which has maximal b'.

Let $\mathcal{B}_1(b') \subset \mathcal{B}$ be the set of all elements of $\mathcal{B}$ for which b' is prefix, that is, $b_1 \in \mathcal{B}_1(b')$ implies $b_1 = b' \cdot c$, $p(c) > P(b')$.

In particular $b \in \mathcal{B}_1(b')$ and $b = b' \cdot P(b)$. We claim that $c \in \mathcal{P}$.

Indeed, assume $b_1 = b' \cdot c$ and $c \notin \mathcal{P}$. Then

$$b_1' = \frac{b_1}{P(b_1)} > b'$$

and (3.8) also holds for $b_1 \in \mathcal{B}$ and b'_1, a contradiction to the maximality of b.

Consider

$$\mathcal{B}_2 = \big(\mathcal{B} \smallsetminus \mathcal{B}_1(b')\big) \cup \{b'\}.$$

We have that $\mathcal{B}_2$ is prefix–free and, since f is multiplicative, that

$$\sum_{b \in \mathcal{B}_1(b')} f(b) \le f(b') \cdot \sum_{P(b') < p \le \frac{n}{b'}} f(p) < f(b') \text{ (by assumption (3.8))}$$

and consequently

$$\sum_{b \in \mathcal{B}_2} f(b) > \sum_{b \in \mathcal{B}} f(b),$$

a contradiction to $\mathcal{B} \in \mathcal{L}_f(n)$.

Hence $\mathcal{B} = \mathcal{A} \in \mathcal{L}_f(n)$.

Define now for any positive, multiplicative function f

$$S_f(n) = \max_{\text{suffix–free } B \subset \mathbb{N}^*(n)} \sum_{b \in B} f(b).$$

Basic Lemma 2. *Write*

$$\mathcal{B} = \left\{ \begin{array}{ll} b \in \mathbb{N}^*(n) : & \text{(i)} \displaystyle\sum_{p < \min\left\{\frac{n+1}{b}, P(b)\right\}} f(p) < 1 \textit{ and} \\ & \text{(ii)} \displaystyle\sum_{p < \min\left\{\frac{n+1}{b'}, P(b')\right\}} f(p) \ge 1, \textit{ where } b' = \frac{b}{P(b)} \end{array} \right\}.$$

We assume that (i) always holds if $\min\left\{\frac{n+1}{b}, P(b)\right\} \le 2$ *and that (ii) holds if* $b \in \mathcal{P}$.
Then B is suffix–free and $\sum_{b \in B} f(b) = S_f(n)$.

Since the proof is almost identical with the one given for Basic Lemma 1, we do not present it here.

4 Prefix–free sets: proofs of Theorem 1, 2

Proof of Theorem 1: This case concerns maximal cardinalities $G(n)$. Notice that $G(n) = L_f(n)$, if f is the constant function with value 1. Furthermore, we verify that the set $\mathcal{B}$ in Theorem 1 equals the set $\mathcal{A}$ in the Basic Lemma 1, which implies the result.

Proof of Theorem 2: Now we apply the Basic Lemma 1 to the multiplicative function f defined by

$$f(m) = \frac{1}{m} \text{ for } m \in \mathbb{N}^*(n).$$

Then

$$E(n) = L_f(n)$$

and the set $\mathcal{A}$ has the properties claimed.

Moreover, the uniqueness can be seen from the proof of the Basic Lemma 1 by observing that we cannot have equality in (3.5) and consequently in (3.7), because $\sum\limits_{p\in\mathcal{P}_1} \frac{1}{p}$ is never an integer for any set $\mathcal{P}_1$ of primes.

To prove the lower bound we consider the set

$$\mathcal{A}' = \left\{ a \in \mathbb{N}^*(n) : P(a) > n^{\frac{1}{1+e}+\varepsilon} \text{ and } \frac{a}{P(a)} < n^{\frac{1}{1+e}-\varepsilon} \right\}.$$

By

$$\sum_{p\le x} \frac{1}{p} = \log\log x + c_5 + O(1)$$

we have for every $a \in \mathcal{A}'$

$$\sum_{P(a)<p\le \frac{n}{a}} \frac{1}{p} < \sum_{n^{\frac{1}{1+e}+\varepsilon}<p\le n^{\frac{e}{1+e}+\varepsilon}} \frac{1}{p} < 1$$

if $n > n_5(\varepsilon)$.

Similarly we have for $a' = \frac{a}{P(a)}$

$$\sum_{P(a')<p\le \frac{n}{a'}} \frac{1}{p} > 1.$$

Therefore $\mathcal{A}' \subset \mathcal{A}$, where $\mathcal{A}$ is the set defined in the Basic Lemma 1. Hence $\mathcal{A}'$ is a prefix–free set.

We have

$$\sum_{a\in\mathcal{A}'} \frac{1}{a} = \sum_{\substack{p>n^{\frac{1}{1+e}+\varepsilon} \\ b<n^{\frac{1}{1+e}-\varepsilon} \\ p\cdot b\le n \\ b\in\mathbb{N}^*}} \frac{1}{b\cdot p} = \sum_{p>n^{\frac{1}{1+e}+\varepsilon}} \frac{1}{p} \cdot \left(\sum_{\substack{b<n^{\frac{1}{1+e}-\varepsilon} \\ b<\frac{n}{p} \\ b\in\mathbb{N}^*}} \frac{1}{b} \right) \ge$$

$$\sum_{n^{\frac{e}{1+e}+\varepsilon}>p>n^{\frac{1}{1+e}+\varepsilon}} \frac{1}{p} \cdot \sum_{\substack{b<n^{\frac{1}{1+e}-\varepsilon} \\ b\in\mathbb{N}^*}} \frac{1}{b} \sim \frac{6}{\pi^2} \log n^{\frac{1}{1+e}-\varepsilon} \cdot \sum_{n^{\frac{e}{1+e}+\varepsilon}>p>n^{\frac{1}{1+e}+\varepsilon}} \frac{1}{p} \sim \frac{6}{\pi^2} \log n^{\frac{1}{1+e}-\varepsilon}.$$

Hence

$$\frac{\sum\limits_{a\in\mathcal{A}'}\frac{1}{a}}{\sum\limits_{a\in\mathbb{N}^*(n)}\frac{1}{a}} > \frac{1}{1+e} - \varepsilon \sim 0,2689 - \varepsilon$$

and this proves the lower bound.

To show the upper bound we consider the set

$$\mathcal{A}'' = \mathbb{N}^*\left(n^{\frac{1}{1+e}-\varepsilon}\right).$$

For every element $a \in \mathcal{A}''$ we have

$$\sum_{P(a)<p\leq\frac{n}{a}} \frac{1}{p} > \sum_{n^{\frac{1}{1+e}-\varepsilon}<p\leq n^{\frac{e}{1+e}+\varepsilon}} \frac{1}{p} > 1.$$

Therefore $\mathcal{A}'' \cap \mathcal{A} = \varnothing$, where $\mathcal{A}$ is the set defined in Theorem 2.

Since

$$\sum_{a\in\mathcal{A}''}\frac{1}{a} = \sum_{a\leq n^{\frac{1}{1+e}-\varepsilon}}\frac{1}{a} \sim \frac{6}{\pi^2}\log n^{\frac{1}{1+e}-\varepsilon}$$

we get

$$\frac{\sum\limits_{b\in\mathcal{A}}\frac{1}{b}}{\sum\limits_{b\in\mathbb{N}^*(n)}\frac{1}{b}} = \frac{E(n)}{\sum\limits_{b\in\mathbb{N}^*(n)}\frac{1}{b}} < \frac{e}{1+e} + \varepsilon \sim 0,7311 + \varepsilon.$$

Remark: We are sure that by more detailed consideration of the set $\mathcal{A}$ one can get much better estimates. However to tighten the gap between upper and lower bounds to, say 0.1, seems difficult.

A proof of Proposition 1 can be given directly and easily with the Basic Lemma 1.

5 Sketch of the proof of Theorem 3

If $\mathcal{A}$ is an infinite prefix–free set and for every finite subset $\mathcal{A}'$ of $\mathcal{A}$ we have

$$\sum_a \frac{1}{a\log a} \leq c, \tag{5.1}$$

where the summation is extended over all $a \in \mathcal{A}'$, then (5.1) also holds if the summation is extended over all $a \in \mathcal{A}$. Thus we may assume that $\mathcal{A}$ is finite.

Let x be large enough in terms of the greatest element of $\mathcal{A}$ (and later we will take $x \to \infty$). Consider the sum

$$S \stackrel{\text{def}}{=} \sum_{a \in \mathcal{A}} \sum_{\substack{q \leq x/a \\ p(q) > a}} \frac{1}{aq}. \tag{5.2}$$

Since $\mathcal{A}$ is prefix–free, thus $aq = a'q'$, $a \in \mathcal{A}$, $a' \in \mathcal{A}$, $q \leq x/q$, $q' \leq x/q'$, $p(q) > a$, $p(q') > a'$ implies that $a = a'$, $q = q'$. In other words, the denominators aq in (5.2) are distinct, each of them is $\leq x$, so that for $x \to \infty$ we have

$$S \leq \sum_{n \leq x} \frac{1}{n} = \big(1 + o(1)\big) \log x \quad (\text{as} \quad x \to \infty). \tag{5.3}$$

On the other hand, we have

$$S = \sum_{a \in \mathcal{A}} \frac{1}{a} \sum_{\substack{q \leq x/a \\ p(q) > a}} \frac{1}{q}. \tag{5.4}$$

Since by Mertens' theorem we have

$$\prod_{p \leq y} \left(1 - \frac{1}{p}\right) \sim \frac{c_5}{\log y}, \tag{5.5}$$

thus by using an elementary sieving process, for $x \to \infty$ we may estimate the inner sum in (5.4) in the following way:

$$\begin{aligned}
\sum_{\substack{q \leq x/a \\ p(q) > a}} \frac{1}{q} &= \sum_{d \mid \prod\limits_{p \leq a} p} \mu(d) \sum_{\substack{q \leq x/a \\ d \mid q}} \frac{1}{q} \\
&= \sum_{d \mid \prod\limits_{p \leq a} p} \mu(d) \sum_{t \leq x/ad} \frac{1}{dt} = \sum_{d \mid \prod\limits_{p \leq a} p} \frac{\mu(d)}{d} \sum_{t \leq x/ad} \frac{1}{t} \\
&= \big(1 + o(1)\big) \sum_{d \mid \prod\limits_{p \leq a} p} \frac{\mu(d)}{d} \log x \\
&= \big(1 + o(1)\big) \log x \prod_{p \leq a} \left(1 - \frac{1}{p}\right) = \big(1 + o(1)\big) c_5 \frac{\log x}{\log a}.
\end{aligned} \tag{5.6}$$

By (5.4) and (5.6) we have

$$S = \big(1 + o(1)\big) c_5 \log x \sum_{a \in \mathcal{A}} \frac{1}{a \log a} \quad (\text{as} \quad x \to \infty). \tag{5.7}$$

Now the desired bound follows from (5.3) and (5.7).

Note that we did not use the fact that the a's are square–free so that, extending the notion of prefix to non–squarefree integers, the result could be extended to this more general case as well.

6 Proof of Theorem 5

(i) Let $\mathcal{A}_1 = \mathcal{A}$, and for $j > 1$ let $\mathcal{A}_j$ denote the set of the integers a such that $a \in \mathcal{A}$ and there is a prefix chain of length j in $\mathcal{A}$ whose last element is a:

$$a_{i_1}|_p a_{i_2}|_p \dots |_p a_{i_{j-1}}|_p a, a_{i_1} \in \mathcal{A}, \dots, a_{i_{j-1}} \in \mathcal{A}.$$

We will show by induction on j that if (2.8) and (2.9) hold, and $1 \le j \le k$ (where k is defined by (2.10)), then

$$E(\mathcal{A}_j) = \sum_{a \in \mathcal{A}_j} \frac{1}{a \log a} \begin{cases} = E(\mathcal{A}) - (j-1)c_3 & \text{for } j = 1 \\ > E(\mathcal{A}) - (j-1)c_3 & \text{for } j > 1. \end{cases} \tag{6.1}$$

Indeed, this is trivial for $j = 1$ since then we have $E(\mathcal{A})$ on both sides of (6.1). Assume now that (6.1) holds for some j with $1 \le j < k$. Then we have to show that it also holds with $j+1$ in place of j:

$$E(\mathcal{A}_{j+1}) = \sum_{a \in \mathcal{A}_{j+1}} \frac{1}{a \log a} > E(\mathcal{A}) - jc_3. \tag{6.2}$$

We will prove this by contradiction: assume that contrary to (6.2) we have

$$E(\mathcal{A}_{j+1}) = \sum_{a \in \mathcal{A}_{j+1}} \frac{1}{a \log a} \le E(\mathcal{A}) - jc_3. \tag{6.3}$$

Write $\mathcal{A}^* = \mathcal{A}_j \setminus \mathcal{A}_{j+1}$. Then by (6.1) and (6.3) (and since clearly $\mathcal{A}_{j+1} \subset \mathcal{A}_j$) we have

$$E(\mathcal{A}^*) = \sum_{a \in \mathcal{A}^*} \frac{1}{a \log a} = E(\mathcal{A}_j) - E(\mathcal{A}_{j+1}) \ge \big(E(\mathcal{A}) - (j-1)c_3\big) - \big(E(\mathcal{A}) - jc_3\big) = c_3.$$

Thus by Theorem 3 there are $a' \in \mathcal{A}^*$, $a'' \in \mathcal{A}^*$ with $a'|_p a''$. Since $a' \in \mathcal{A}^* \subset \mathcal{A}_j$, thus there is a prefix chain of length j in $\mathcal{A}$ whose last element is $a' : a_{i_1}|_p \dots |_p a'$. Then $a_{i_1}|_p \dots |_p a_{i_{j-1}}|_p a'|_p a''$ is a prefix chain of length $j+1$ in $\mathcal{A}$ whose last element is a'', and thus we have $a'' \in \mathcal{A}_{j+1}$. This contradicts $a'' \in \mathcal{A}^* = \mathcal{A}_j \setminus \mathcal{A}_{j+1}$ which proves (6.2), and this completes the proof of (6.1) (with $1 \le j \le k$).

Using (6.1) with k in place of j (so that $k \ge 2$ by (2.9)) we obtain

$$E(\mathcal{A}_k) > E(\mathcal{A}) - (k-1)c_3 = E(\mathcal{A}) - \left[\frac{E(\mathcal{A})}{c_3}\right] c_3 \ge O$$

so that $\mathcal{A}_k$ is non–empty, which completes the proof of (i).

(ii) Let

$$\mathcal{B} = \{b : b \in \mathbb{N}, |\omega(b) - \log\log b| < (\log\log b)^{3/4}\}$$

and $\mathcal{A} = \mathcal{B} \cap \mathbb{N}^*$. Then by a theorem of Hardy and Ramanujan [13] we have $d(\mathcal{B}) = 1$, which implies (2.11). Moreover, by (1.1) clearly we have

$$E(\mathcal{A}, n) = \sum_{a \in \mathcal{A}, a \le n} \frac{1}{a \log a} = \big(1 + o(1)\big) \sum_{a \in \mathbb{N}^*(n)} \frac{1}{a \log a} = \big(1 + o(1)\big) \frac{6}{\pi^2} \log\log n. \tag{6.4}$$

If $a \in \mathcal{A}$, $a \le n$ and a is the last element of a prefix chain of length k in $\mathcal{A} : a_{i_1}|_p \dots |_p a_{i_{k-1}}|_p a$, then by (6.4) and the definition of $\mathcal{A}$ we have

$$k \le \omega(a) < \log\log a + (\log\log a)^{3/4} \le \log\log n + (\log\log n)^{3/4}$$

$$= \big(1 + o(1)\big) \log\log n = \big(1 + o(1)\big) \frac{\pi^2}{6} E(\mathcal{A}, n),$$

which completes the proof of (ii) (with $c_4 = \frac{\pi^2}{6} + \varepsilon$).

7 Proof of Theorem 6

We apply Basic Lemma 2 with respect to the function $f(m) = 1$, $m \in \mathbb{N}^*$. For this function we have $H(n) = S_f(n)$. It is easy to verify, that $\mathcal{C}(n) \subset \mathcal{B}$, where $\mathcal{C}(n)$ is the set described in Theorem 6 and $\mathcal{B}$ is the set from Basic Lemma 2. Moreover, for every $a \in \mathbb{N}^*(n) \smallsetminus \mathcal{C}(n)$ we have $2 \nmid a$, $a \le \frac{n}{2}$ and hence

$$\min\left\{\frac{n+1}{a}, P(a)\right\} > 2.$$

Consequently, the condition (i) in Basic Lemma 2 does not hold. Therefore $\mathcal{C}(n) = \mathcal{B}$ and $\mathcal{C}(n)$ is the optimal set.

8 Proof of Theorem 7

In Basic Lemma 2 consider the set $\mathcal{B}$ with respect to the multiplicative function $f(m) = \frac{1}{m}$, $m \in \mathbb{N}^*$. Using the inequalities $\frac{1}{2} + \frac{1}{3} = \frac{5}{6} < 1$ and $\frac{1}{2} + \frac{1}{3} + \frac{1}{5} = \frac{31}{30} > 1$, it is easy to verify that $(\mathcal{B}_1 \dot\cup \mathcal{B}_2) \subset \mathcal{B}$, where $\mathcal{B}_1, \mathcal{B}_2$ are defined in the Theorem. Moreover, using the mentioned inequalities one easily gets that every $b \in \mathbb{N}^*(n) \smallsetminus \{\mathcal{B}_1 \cup \mathcal{B}_2)$ violates one of the conditions (i), (ii) in Basic Lemma 2.

Hence $\mathcal{B} = \mathcal{B}_1 \dot\cup \mathcal{B}_2$, proving the Theorem.

Corollary 6 and 7 directly follow from Theorem 7 and from the construction.

Finally, Proposition 2 is an immediate consequence of Basic Lemma 2.

References

[1] R. Ahlswede, L. Khachatrian and A. Sárközy, On the counting function of primitive sets of integers, Preprint 98–077, SFB 343 “Diskrete Strukturen in der Mathematik”, Universität Bielefeld, submitted to J. Number Theory.

[2] R. Ahlswede and L.H. Khachatrian, Classical results on primitive and recent results on cross–primitive sequences, in: The Mathematics of Paul Erdös, vol. I, eds.R.L. Graham and J. Nesetril, Algorithms and Combinatorics 13, Springer–Verlag, 1997; 104–116.

[3] A.S. Besicovitch, On the density of certain sequences, Math. Ann. 110 (1934), 336–341.

[4] F. Behrend, On sequences of numbers not divisible by one another, J. London Math. Soc. 10 (1935), 42–44.

[5] P. Erdös, Note on sequences of integers no one of which is divisible by any other, J. London Math. Soc. 10 (1935), 126–128.

[6] P. Erdös, A generalization of a theorem of Besicovitch, J. London Math. Soc. 11 (1935), 92–98.

[7] P. Erdös, A. Sárközy and E. Szemerédi, On a theorem of Behrend, J. Australian Math. Soc. 7 (1967), 9–16.

[8] P. Erdös, A. Sárközy and E. Szemerédi, On divisibility properties of sequences of integers, Coll. Math. Soc. J. Bolyai 2 (1970), 35–49.

[9] H. Halberstam and K.F. Roth, Sequences, Springer–Verlag, Berlin–Heidelberg–New York, 1983.

[10] A. Sárközy, On divisibility properties of sequences of integers, in: The Mathematics of Paul Erdös, eds. R.L. Graham and J. Nesetril, Algorithms and Combinatorics 13, Springer–Verlag, 1997; 241–250.

[11] A. Selberg, Note on a paper by L.G. Sathe, J. Indian Math. Soc. 18 (1954), 83–87.

[12] H. Davenport and P. Erdös, On sequences of positive integers, Acta Arith. 2 (1936), 147–151.

[13] G.H. Hardy and S. Ramanujan, The normal number of prime factors of a number n, Quarterly J. Math. 48 (1920), 76–92.