On primitive sets of squarefree integers

R. Ahlswede and L. Khachatrian
Fakultät für Mathematik
Universität Bielefeld
Postfach 100131
33501 Bielefeld

and

A. Sárközy *
Eötvös Loránd University
Department of Algebra and Number Theory
H–1088 Budapest, Rákóczi út 5
Hungary

* Research partially supported by Hungarian National Foundation for Scientific Research, Grant No. T029759.
This paper was written while A. Sárközy was visiting the Universität Bielefeld.

1 Introduction

The set of positive integers is denoted by $\mathbb{N}$ while the set of positive squarefree integers is denoted by $\mathbb{N}^*$. For $\mathcal{A} \subset \mathbb{N}$ we write

$$S(\mathcal{A}) = \sum_{a \in \mathcal{A}} \frac{1}{a}.$$

A set $\mathcal{A} \subset \mathbb{N}$ is said to be primitive, if there are no $a \in \mathcal{A}$, $a' \in \mathcal{A}$ with $a \neq a'$, $a|a'$. The family of the primitive sets $\mathcal{A} \subset \mathbb{N}$ is denoted by $\mathbb{P}$, and $\mathbb{P}^*$ denotes the family of the primitive sets $\mathcal{A}$ consisting of squarefree integers, i.e., with $\mathcal{A} \in \mathbb{P}$, $\mathcal{A} \subset \mathbb{N}^*$. A subscript N indicates if we restrict ourselves to integers not exceeding N, so that $\mathbb{N}_N = \{1, 2, \ldots, N\}$; $\mathbb{N}^*_N$ denotes the set of the (positive) squarefree integers not exceeding N; $\mathbb{P}_N$ is the family of the primitive subsets of $\{1, 2, \ldots, N\}$; $\mathbb{P}^*_N$ denotes the family of the primitive sets selected from the squarefree integers not exceeding N. The number of distinct prime factors of n is denoted by $w(n)$, while $\Omega(n)$ denotes the total number (counted with multiplicity) of prime factors of n:

$$\Omega(n) = \sum_{p^\alpha \| n} \alpha.$$

(Here $p^\alpha \| n$ denotes that $p^\alpha | n$ but $p^{\alpha+1} \nmid n$.)

It is well–known and easy to prove (see, e.g., [8, p. 244]) that

$$\max_{\mathcal{A} \in \mathbb{P}_N} |\mathcal{A}| = N - [N/2] \quad \left(= \left(\frac{1}{2} + o(1) \right) N \right). \tag{1.1}$$

Behrend [2] proved that

$$\max_{\mathcal{A} \in \mathbb{P}_N} S(\mathcal{A}) < c_1 \frac{\log N}{(\log \log N)^{1/2}}$$

for some absolute constant c_1 and all $N \geq 3$ and Erdős, Sárközy and Szemerédi [6] determined the value of the best possible constant in the following sense:

$$\max_{\mathcal{A} \in \mathbb{P}_N} S(\mathcal{A}) = \big(1 + o(1)\big) \frac{\log N}{(2\pi \log \log N)^{1/2}} \text{ as } N \to \infty. \tag{1.2}$$

Erdős [3] proved that

$$\sum_{a \in \mathcal{A}} \frac{1}{a \log a} < c_2 \text{ for all } \mathcal{A} \in \mathbb{P} \text{ with } 1 \notin \mathcal{A} \tag{1.3}$$

where c_2 is an absolute constant, and he conjectured (see [13]) that the sum in (1.3) is maximal if $\mathcal{A}$ is the set of the primes:

$$\sum_{a \in \mathcal{A}} \frac{1}{a \log a} \leq \sum_{p} \frac{1}{p \log p} \text{ for all } \mathcal{A} \in \mathbb{P} \text{ with } 1 \notin \mathcal{A}. \tag{1.4}$$

This conjecture has not been proved yet; Erdős and Zhang have several partial results.

In this paper our goal is to study the squarefree analogs of the problems in (1.1) and (1.2), i.e., to estimate

$$\max_{\mathcal{A}\in\mathbb{P}^*_N} |\mathcal{A}| \quad \text{and} \quad \max_{\mathcal{A}\in\mathbb{P}^*_N} S(\mathcal{A}). \tag{1.5}$$

2 The results

Since

$$\lim_{N\to\infty} |\mathbb{N}^*_N|/|\mathbb{N}_N| = \frac{1}{\zeta(2)} = \frac{6}{\pi^2},$$

i.e., roughly speaking the proportion of integers being squarefree is $\frac{6}{\pi^2}$, thus one expects that the maxima in (1.5) are less than the ones in (1.1) and (1.2) by a factor $\frac{6}{\pi^2}+o(1)$. In case of the second maximum these rough heuristics lead to the following conjecture:

Conjecture 1. *We have*

$$\max_{\mathcal{A}\in\mathbb{P}^*_N} S(\mathcal{A}) = \big(1+o(1)\big)\frac{6}{\pi^2}\frac{\log N}{(2\pi\log\log N)^{1/2}} \ \textit{as } N\to\infty. \tag{2.1}$$

This conjecture was raised by Pomerance and Sárközy in [10]. Indeed, we will prove this conjecture in the following more general form:

Theorem 1. *Let $Q=\{q_1,q_2,\dots\}=\{p_1^{\alpha_1},p_2^{\alpha_2},\dots\}$ (with $p_1<p_2<\dots$) be a set of powers of distinct primes with*

$$S(Q)<\infty. \tag{2.2}$$

Then we have

$$\max_{\substack{\mathcal{A}\in\mathbb{P}_N\\ q\nmid a \text{ for } a\in\mathcal{A},q\in Q}} S(\mathcal{A}) = \big(1+o(1)\big)\prod_{q\in Q}\left(1-\frac{1}{q}\right)\frac{\log N}{(2\pi\log\log N)^{1/2}} \ \textit{as } N\to\infty \tag{2.3}$$

Note that here $Q=\varnothing$ is allowed and, indeed, in this special case we obtain theorem (1.2) of Erdős, Sárközy and Szemerédi.

Choosing

$$Q=\{2^2,3^2,5^2,\dots,p^2,\dots\}, \tag{2.4}$$

we obtain

Corollary 1. *(2.1) holds, i.e., Conjecture 1 is true.*

Another important special case is when Q consists of the primes not exceeding a fixed number K:

Corollary 2. *If $K \geq 2$, then we have*

$$\max_{\substack{\mathcal{A}\in\mathbb{P}_N \\ \left(a,\prod_{p\leq K} p\right)=1 \text{ for all } a\in\mathcal{A}}} S(\mathcal{A}) = \big(1+o(1)\big)\prod_{p\leq K}\left(1-\frac{1}{p}\right)\frac{\log N}{(2\pi\log\log N)^{1/2}} \quad \text{as } N\to\infty.$$

Moreover, we can prove that if Q is finite, then Q need not consist of prime powers, it suffices to assume coprimality:

Theorem 1'. *Let $Q = \{q_1,\dots,q_t\}$ be a finite set of pairwise coprime positive integers: $(q_i,q_j)=1$ for $1\leq i<j\leq t$. Then (2.3) holds.*

Since the proof of Theorems 1 and 1' are similar and, indeed, the proof of the latter is slightly easier, we will prove only Theorem 1 here. (On the other hand, the proof of a general theorem, which includes both theorems as special cases, could be more troublesome.)

It comes perhaps as a surprise that the heuristics at the beginning of Section 2 fails in case of the first maximum in (1.5). Indeed, an integer $m\in\mathbb{N}_N$ (resp. $\mathbb{N}_N^*$) is called maximal in $\mathbb{N}_N$ (resp. $\mathbb{N}_N^*$) if $m\nmid m'$ for all $m\neq m'$, $m'\in\mathbb{N}_N$ (resp. $\mathbb{N}_N^*$). Let $M(N)$ (resp. $M^*(N)$) denote the set of all maximal integers in $\mathbb{N}_N$ (resp. $\mathbb{N}_N^*$). Clearly

$$M(N)\in\mathbb{P}_N, M^*(N)\in\mathbb{P}_N^* \quad \text{and}$$

$$\begin{aligned}
M(N) &= \left\{n\in\mathbb{N}_N : \frac{N}{2}<n\leq N\right\} \quad \text{and} \\
M^*(N) &= \left\{n\in\mathbb{N}_N^* : \frac{N}{2}<n\leq N\right\}\cup\left\{n\in\mathbb{N}_N^* : 2\mid n \text{ and } \frac{N}{3}<n\leq\frac{N}{2}\right\}\cup \\
&\quad \left\{n\in N_N^* : 6\mid n \text{ and } \frac{N}{5}<n\leq\frac{N}{3}\right\}\cup \\
&\quad \left\{n\in N_N^* : 2\cdot3\cdot5\dots p_k\mid n \text{ and } \frac{N}{p_{k+1}}<n\leq\frac{N}{p_k}\right\}\cup\dots
\end{aligned}$$

(p_k is the k–th prime).

Hence $|M(N)| = N-\left\lfloor\frac{N}{2}\right\rfloor = \max_{\mathcal{A}\in\mathbb{P}_N}|\mathcal{A}|$ (by (1.1)) and simple calculation shows that

$$|M^*(N)| = \left(\frac{2}{3}-\sum_{k\geq2}\frac{1}{(p_1+1)(p_2+1)\dots(p_k+1)}+o(1)\right)\cdot N_N^* \sim (0,5676+o(1))\cdot N_N^* \leq \max_{\mathcal{A}\in\mathbb{P}_N^*}|\mathcal{A}|.$$

We will prove that this maximum is much greater but can be estimated surprisingly well.

Theorem 2. *For $N>N_o$ we have*

(i) $\max_{\mathcal{A}\in\mathbb{P}_N^*}|\mathcal{A}| > 0,6328\frac{6}{\pi^2}N.$

(ii) $\max_{\mathcal{A}\in\mathbb{P}_N^*} |\mathcal{A}| < 0,6389\frac{6}{\pi^2}N$.

(Actually, we can show by the same method but with more computation that for large N we have $0,6362\frac{6}{\pi^2}N < \max_{\mathcal{A}\in\mathbb{P}_N^*} |\mathcal{A}| < 0,6366\frac{6}{\pi^2}N$.)

More generally, one might like to count the elements a of $\mathcal{A}$ with certain weights $f(a)$ where $f(x)$ is a "nice" function (we will return to this question in a subsequent paper). Theorem 1 corresponds to the weight $f(a) = \frac{1}{a}$, while in case of Theorem 2 the weighting is $f(a) = 1$ (for all $a \in \mathbb{N}$). If Erdős's conjecture (1.4) is right, then the heuristics above also fails in case of the weighting $f(a) = \frac{1}{a\log a}$, since then we have

$$\max_{\substack{\mathcal{A}\in\mathbb{P}\\ 1\notin\mathcal{A}}} \sum_{a\in\mathcal{A}} \frac{1}{a\log a} = \max_{\substack{\mathcal{A}\in\mathbb{P}^*\\ 1\notin\mathcal{A}}} \sum_{a\in\mathcal{A}} \frac{1}{a\log a} = \sum_{p} \frac{1}{p\log p}.$$

One might like to study what happens for other weightings. Another important special case is $f(a) = \frac{1}{a^\sigma}$ with $0 < \sigma < 1$ ($\sigma = 0$ and $\sigma = 1$ correspond to the special cases $f(a) = 1$, resp. $f(a) = \frac{1}{a}$ studied above).

We introduce for $0 \le \sigma \le 1$

$$F(\sigma) = \lim_{N\to\infty} \left(\max_{\mathcal{A}\in\mathbb{P}_N^*} \sum_{a\in\mathcal{A}} \frac{1}{a^\sigma} \right) \Big/ \left(\max_{\mathcal{A}\in\mathbb{P}_N} \sum_{a\in\mathcal{A}} \frac{1}{a^\sigma} \right),$$

if the limit exists and conjecture that this is the case for $\sigma = 0$. Then by Theorem 2 $F(0) > \frac{6}{\pi^2}$. Moreover we have the following:

Conjecture 2.

(i) *For all* $0 < \sigma < 1$ *the limit* $F(\sigma)$ *exists.*

(ii) $F(\sigma) > \frac{6}{\pi^2}$.

(iii) $\lim_{\sigma\to 1^-} F(\sigma) = \frac{6}{\pi^2}$.

(iv) $F(\sigma)$ *is decreasing in* $(0,1)$.

Finally, we draw attention to our paper [1], where we studied normalized weights in the cases $\sigma = 0, 1$ for prefix–free (and also suffix–free) sets of numbers rather than primitive sets and obtained several conclusive results. In Theorem 2 there the function

$$E(N) = \left(\max_{\text{prefix–free } \mathcal{A}\subset\mathbb{N}_N^*} \sum_{a\in\mathcal{A}} \frac{1}{a} \right) \Big/ \left(\sum_{a\in\mathbb{N}_N^*} \frac{1}{a} \right)$$

is bounded by constants from below and above for all large N. Here $\lim_{N\to\infty} E(N)$ is also conjectured to exist. If so, what is its value?

3 Auxiliary results

In order to prove Theorem 1, we need several lemmas.

Lemma 1. *For $x \to \infty$ we have*

$$\sum_{p \le x} \frac{1}{p} = \log\log x + c_3 + O\left(\frac{1}{\log x}\right)$$

and

$$\sum_{p^\alpha \le x} \frac{1}{p^\alpha} = \log\log x + c_4 + O\left(\frac{1}{\log x}\right).$$

(In the second sum the summation is over all prime powers not exceeding x.)

These two (equivalent) formulas can be found in any book on prime number theory (see, e.g., [9, p. 20]).

Lemma 2. *Write*

$$\mathcal{P}(x,k) = \big\{n : n \le x, \Omega(n) = k\big\}. \tag{3.1}$$

Then uniformly for $x \ge 3$, $k \in \mathbb{N}$ we have

$$\sum_{n \in \mathcal{P}(x,k)} \frac{1}{n} < c_5 \frac{\log x}{(\log\log x)^{1/2}}.$$

Proof: This follows from [4] and [5] (see also [11] and [12]) and as pointed out by the referee also directly from Lemma 1 and the inequality $\frac{a^k}{k!} \ll \frac{e^a}{a^{1/2}}$, uniformly for all $a > 1$ and all $k \in \mathbb{N}$.

Lemma 3. *For any f with $f(x) > 0$ for $x > x_0$ and*

$$\lim_{x \to \infty} f(x)(\log\log x)^{-1/2} = 0,$$

we have uniformly for $k \in \mathbb{N}$, satisfying

$$\log\log x - f(x) < k < \log\log x + f(x), \tag{3.2}$$

$$\sum_{n \in \mathcal{P}(x,k)} \frac{1}{n} = \big(1 + o(1)\big) \frac{\log x}{(2\pi \log\log x)^{1/2}} \ \text{ as } x \to \infty$$

(where $\mathcal{P}(x,k)$ is defined by (3.1)).

Proof: See [4].

Lemma 4. *Define $Q = \{q_1, q_2, \dots\}$ as in Theorem 1 (so that (2.2) holds), write*

$$\mathcal{B}(x,k,Q) = \big\{n : n \in \mathcal{P}(x,k), q \nmid n \text{ for } q \in Q\big\}, \tag{3.3}$$

and define $f(x)$ as in Lemma 3. Then uniformly for k satisfying (3.2) we have

$$\sum_{n\in\mathcal{B}(x,k,Q)}\frac{1}{n}=\big(1+o(1)\big)\prod_{q\in Q}\left(1-\frac{1}{q}\right)\frac{\log x}{(2\pi\log\log x)^{1/2}}\ \textit{as } x\to\infty. \tag{3.4}$$

Proof: Fix an $\varepsilon>0$, and let H be a positive integer with

$$\sum_{\substack{q\in Q\\ q>H}}\frac{1}{q}<\varepsilon; \tag{3.5}$$

by (2.2), such a number H exists. Write

$$Q_1=\{q:q\in Q,q\le H\}$$

and

$$Q_2=\{q:q\in Q,H<q\le x\}.$$

Clearly we have

$$\mathcal{B}(x,k,Q)\subset\mathcal{B}(x,k,Q_1),$$

and if $n\in\mathcal{B}(x,k,Q_1)$ but $n\notin\mathcal{B}(x,k,Q)$, then we have $n\in\mathcal{P}(x,k)$ and $q|n$ for some $q\in Q_2$. It follows that

$$\left|\sum_{n\in\mathcal{B}(x,k,Q)}\frac{1}{n}-\sum_{n\in\mathcal{B}(x,k,Q_1)}\frac{1}{n}\right|\le\sum_{q\in Q_2}\sum_{\substack{n\in\mathcal{P}(x,k)\\ q|n}}\frac{1}{n}. \tag{3.6}$$

By the exclusion–inclusion principle we have

$$\begin{aligned}\sum_{n\in\mathcal{B}(x,k,Q_1)}\frac{1}{n}&=\sum_{n\in\mathcal{P}(x,k)}\frac{1}{n}+\sum_{j}(-1)^j\sum_{\substack{q_{i_1},\dots,q_{i_j}\in Q_1\\ q_{i_1}<\dots<q_{i_j}}}\ \sum_{\substack{n\in\mathcal{P}(x,k)\\ q_{i_1}\cdots q_{i_j}|n}}\frac{1}{n}\\ &=\sum_{n\in\mathcal{P}(x,k)}\frac{1}{n}+\sum_{j}(-1)^j\sum_{q_{i_1}<\dots<q_{i_j}\le H}\frac{1}{q_{i_1}\dots q_{i_j}}\sum_{t\in\mathcal{P}(x/q_{i_1}\cdots q_{i_j},k-\Omega(q_{i_1}\cdots q_{i_j})}\frac{1}{t}.\end{aligned}$$

By Lemma 3, it follows that for $x\to\infty$ we have

$$\begin{aligned}&\sum_{n\in\mathcal{B}(x,k,Q_1)}\frac{1}{n}\\ &=\big(1+o(1)\big)\left(\frac{\log x}{(2\pi\log\log x)^{1/2}}+\sum_{j}(-1)^j\sum_{q_{i_1}<\dots<q_{i_j}\le H}\frac{1}{q_{i_1}\dots q_{i_j}}\frac{\log(x/q_{i_1}\dots q_{i_j})}{(2\pi\log\log(x/q_{i_1}\dots q_{i_j}))^{1/2}}\right)\\ &=\big(1+o(1)\big)\prod_{\substack{q\in Q\\ q\le H}}\left(1-\frac{1}{q}\right)\frac{\log x}{(2\pi\log\log x)^{1/2}}.\end{aligned} \tag{3.7}$$

By (3.5) here we have

$$\prod_{q\in Q}\left(1-\frac{1}{q}\right)\le\prod_{\substack{q\in Q\\ q\le H}}\left(1-\frac{1}{q}\right)=\prod_{q\in Q}\left(1-\frac{1}{q}\right)\prod_{\substack{q\in Q\\ q>H}}\left(1-\frac{1}{q}\right)^{-1}<\prod_{q\in Q}\left(1-\frac{1}{q}\right)\prod_{\substack{q\in Q\\ q>H}}\left(1+\frac{2}{q}\right)$$

$$<\prod_{q\in Q}\left(1-\frac{1}{q}\right)\exp\left(\sum_{\substack{q\in Q\\ q>H}}\frac{2}{q}\right)<\prod_{q\in Q}\left(1-\frac{1}{q}\right)\exp(2\varepsilon). \tag{3.8}$$

Moreover, by (3.5) and Lemma 2 we have

$$\sum_{q\in Q_2}\sum_{\substack{n\in\mathcal{P}(x,k)\\ q|n}}\frac{1}{n}=\sum_{q\in Q_2}\sum_{qt\in\mathcal{P}(x,k)}\frac{1}{qt}=\sum_{q\in Q_2}\frac{1}{q}\sum_{t\in\mathcal{P}(x/q,k-\Omega(q))}\frac{1}{t}$$

$$<\sum_{q\in Q_2}\frac{1}{q}\,c_5\,\frac{\log x}{(\log\log x)^{1/2}}<c_5\,\varepsilon\,\frac{\log x}{(\log\log x)^{1/2}}. \tag{3.9}$$

Since (3.6), (3.7), (3.8) and (3.9) hold for $\varepsilon>0$ and $x>x_0(\varepsilon)$, thus (3.4) follows.

Lemma 5. *Write*

$$z=[\log\log N]. \tag{3.10}$$

Then we have

$$S\big(\{n:n\le N,\Omega(n)-\omega(n)>100\log z\}\big)=o\left(\frac{\log N}{(\log\log N)^{1/2}}\right).$$

Proof: This is the Lemma in [6].

Lemma 6. *For $n\ge 30$ we have*

$$\sum_{p|n}\frac{1}{p}<c_6\log\log\log n. \tag{3.11}$$

Proof: Clearly we have

$$\prod_{p|n}\left(1+\frac{1}{p}\right)<\prod_{p|n}\frac{1}{1-\frac{1}{p}}=\frac{n}{\varphi(n)}$$

whence, by the well–known inequality

$$\varphi(n)>c_7\,\frac{n}{\log\log n}$$

(see, e.g. [9, p. 24]), it follows that

$$\prod_{p\mid n}\left(1+\frac{1}{p}\right) < c_8 \log\log n. \tag{3.12}$$

On the other hand, we have

$$\begin{aligned}\prod_{p\mid n}\left(1+\frac{1}{p}\right) &= \exp\left(\sum_{p\mid n}\log\left(1+\frac{1}{p}\right)\right)\\ &= \exp\left(\sum_{p\mid n}\frac{1}{p}+O\left(\sum_{p}\frac{1}{p^2}\right)\right) > \exp\left(\sum_{p\mid n}\frac{1}{p}+c_9\right).\end{aligned} \tag{3.13}$$

(3.11) follows from (3.12) and (3.13).

4 Completion of the proof of Theorem 1

The lower bound for the maximum in (2.3) is a straightforward consequence of Lemma 4. Indeed, define again z by (3.10) and let $\mathcal{A}^* = \mathcal{B}(N,z,Q)$ (where $\mathcal{B}(N,z,Q)$ is defined by (3.3)). Then clearly we have

$$\mathcal{A}^* \in \mathbb{P}_N,\ \ q \nmid a \text{ for } a \in \mathcal{A}^*, q \in Q$$

and, by Lemma 4,

$$S(\mathcal{A}^*) = \sum_{n\in\mathcal{B}(N,z,Q)}\frac{1}{n} = \big(1+o(1)\big)\prod_{q\in Q}\left(1-\frac{1}{q}\right)\frac{\log N}{(2\pi\log\log N)^{1/2}}.$$

It follows that the maximum in (2.3) is

$$\max_{\substack{\mathcal{A}\in\mathbb{P}_N\\ q\nmid a \text{ for } a\in\mathcal{A}, q\in Q}} S(\mathcal{A}) \geq S(\mathcal{A}^*) = \big(1+o(1)\big)\prod_{q\in Q}\left(1-\frac{1}{q}\right)\frac{\log N}{(2\pi\log\log N)^{1/2}}. \tag{4.1}$$

In order to prove that this bound is also an upper bound for $\max S(\mathcal{A})$, we will adopt the method of the proof in [6] (which is a method of combinatorial flavour and it reminds one of the proof of Sperner's theorem).

Consider an $\mathcal{A}$ with $\mathcal{A} \in \mathbb{P}_N$ and such that $q \nmid a$ for $a \in \mathcal{A}, q \in Q$. Define $\mathcal{A}'$ by $\mathcal{A}' = \big\{a : a \in \mathcal{A}, \Omega(a)-\omega(a) \leq 100\log z\big\}$ where again, z is defined by (3.10). By (2.2) and Lemma 5, it suffices to prove that

$$S(\mathcal{A}') \leq \big(1+o(1)\big)\prod_{q\in Q}\left(1-\frac{1}{q}\right)\frac{\log N}{(2\pi\log\log N)^{1/2}} \text{ as } N\to\infty. \tag{4.2}$$

Following combinatorial terminology, we will refer to the integers n with $\Omega(n) = k$ as "integers on level k". First we separate the a's of different level, i.e., we write

$$\mathcal{A}_k = \{a : a \in \mathcal{A}', \Omega(a) = k\}.$$

Let r_1 and r_2 denote the occuring smallest, resp. greatest level, so that

$$\mathcal{A}' = \bigcup_{k=r_1}^{r_2} \mathcal{A}_k$$

and

$$S(\mathcal{A}') = \sum_{k=r_1}^{r_2} S(\mathcal{A}_k).$$

Now we separate the contribution of the $\mathcal{A}_k$'s with $k > z$ ("high level a's"; here again z is defined by (3.10)), $k = z$, resp. $k < z$ ("low level a's"):

$$S(\mathcal{A}') = \sum_{z<k\leq r_2} S(\mathcal{A}_k) + S(\mathcal{A}_z) + \sum_{r_1\leq k<z} S(\mathcal{A}_k) = \Sigma_1 + \Sigma_2 + \Sigma_3. \tag{4.3}$$

We will compare Σ_1 and Σ_3 with the sum of the reciprocals of the numbers having z prime factors which can be constructed from the "high level a's" by dropping prime factors, resp. from the "low level a's" by adding prime factors (without producing a multiple of a q). Indeed, let $\mathcal{D}$ denote the set of the integers d such that $\Omega(d) = z$ and there is an $a \in \mathcal{A}_k$ with $d|a$, $k > z$, and let $\mathcal{E}$ denote the set of the integers e such that $\Omega(e) = z$, there is no $q \in Q$ with $q|e$, and there is an $a \in \mathcal{A}_k$ with $a|e$, $k < z$. It follows from the primitivity of $\mathcal{A}'$ that

$$\mathcal{D} \cap \mathcal{E} = \mathcal{D} \cap \mathcal{A}_z = \mathcal{E} \cap \mathcal{A}_z = \varnothing, \tag{4.4}$$

and clearly we have

$$\mathcal{D} \subset \mathcal{B}(N, z, Q), \mathcal{E} \subset \mathcal{B}(N, z, Q) \text{ and } \mathcal{A}_z \subset \mathcal{B}(N, z, Q). \tag{4.5}$$

By Lemma 4, it follows from (4.4) and (4.5) that

$$S(\mathcal{D}) + S(\mathcal{E}) + S(\mathcal{A}_z) \leq S\big(\mathcal{B}(N, z, Q)\big) = \big(1 + o(1)\big) \prod_{q\in Q} \left(1 - \frac{1}{q}\right) \frac{\log N}{(2\pi \log\log N)^{1/2}}. \tag{4.6}$$

We will show that

$$\Sigma_1 \leq \big(1 + o(1)\big) S(\mathcal{D}) \tag{4.7}$$

and

$$\Sigma_3 \leq \big(1 + o(1)\big) S(\mathcal{E}) + o\left(\frac{\log N}{(\log\log N)^{1/2}}\right). \tag{4.8}$$

(4.2) would follow from (4.3), (4.6), (4.7) and (4.8) and this, together with (4.1), would complete the proof of the theorem. Thus it remains to prove (4.7) and (4.8).

First we will prove (4.7). For $z \le k \le r_2$ define the sets $\mathcal{U}_k$, $\mathcal{F}_k$ by the following recursion: Let $\mathcal{U}_{r_2} = \varnothing$ and $\mathcal{F}_{r_2} = \mathcal{A}_{r_2} (= \mathcal{A}_{r_2} \cup \mathcal{U}_{r_2})$. If $z < k \le r_2$ and $\mathcal{U}_k$, $\mathcal{F}_k$ have been defined, then let

$$\mathcal{U}_{k-1} = \big\{ u : u \in \mathbb{N}, \Omega(u) = k-1, \text{ there is } f \in \mathcal{F}_k \text{ with } u|f \big\}$$

and

$$\mathcal{F}_{k-1} = \mathcal{A}_{k-1} \cup \mathcal{U}_{k-1}. \tag{4.9}$$

Then clearly we have

$$\mathcal{D} = \mathcal{U}_z. \tag{4.10}$$

Moreover, by the primitivity of $\mathcal{A}$, for all $z \le k \le r_2$ we have

$$\mathcal{A}_k \cap \mathcal{U}_k = \varnothing. \tag{4.11}$$

If $z \le k < r_2$ and $u \in \mathcal{U}_k$, then there is a number f with $u|f$, $f \in \mathcal{F}_{k+1}$. Then we have $\Omega(u) = k$ and $\Omega(f) = k+1$, so that f is of the form $f = up$. By $f \in \mathcal{F}_{k+1}$ we have $f|a$ for some $a \in \mathcal{A}'$ and thus $\Omega(f) - \omega(f) \le \Omega(a) - \omega(a) \le 100 \log z$. Moreover, again by $f \in \mathcal{F}_{k+1}$ we have $\Omega(f) = k+1$. It follows that f has at least

$$\omega(f) \ge \Omega(f) - 100 \log z = k + 1 - 100 \log z$$

representations in form up so that

$$S(\mathcal{U}_k) \sum_{p \le N} \frac{1}{p} = \sum_{u \in \mathcal{U}_k} \sum_{p \le N} \frac{1}{up} \ge (k + 1 - 100 \log z) \sum_{f \in \mathcal{F}_{k+1}} \frac{1}{f} = (k + 1 - 100 \log z) S(\mathcal{F}_{k+1}).$$

By Lemma 1, (4.9) and (4.11), it follows that

$$S(\mathcal{U}_k) > \frac{k + 1 - 100 \log z}{z + c_{10}} \big(S(\mathcal{A}_{k+1}) + S(\mathcal{U}_{k+1}) \big) \tag{4.12}$$

for $z \le k < r_2$. For N large enough clearly we have

$$\frac{k + 1 - 100 \log z}{z + c_{10}} > \begin{cases} 1 & \text{if } k > z + 200 \log z \\ 1 - \frac{200 \log z}{z} & \text{for all } k \ge z. \end{cases} \tag{4.13}$$

It follows from (4.12) and (4.13) by a simple induction argument that

$$S(\mathcal{U}_z) > \left(1 - \frac{200 \log z}{z} \right)^{200 \log z} \sum_{z < k \le r_2} S(\mathcal{A}_k) = \big(1 + o(1) \big) \Sigma_1$$

which, by (4.10), proves (4.7).

Now we will prove (4.8). The proof is similar to the proof of (4.7), but while there in each step we moved one level lower by dropping a prime factor, here we move one level higher by adding a prime. Moreover, here a little difficulty arises from the facts that, first, we have to restrict ourselves to integers v with $q \nmid v$ for $q \in Q$, and, secondly, also integers v slightly greater than N appear.

For $r_1 \le k \le z$ define the sets $\mathcal{V}_k$, $\mathcal{G}_k$ by the following recursion: Let $\mathcal{V}_{r_1} = \varnothing$ and $\mathcal{G}_{r_1} = \mathcal{A}_{r_1}$ $(= \mathcal{A}_{r_1} \cup \mathcal{V}_{r_1})$. If $r_1 \le k < z$ and $\mathcal{V}_k$, $\mathcal{G}_k$ have been defined, then let $\mathcal{V}_{k+1}$ denote the set of the positive integers v that can be represented in the form $v = pg$ with $g \in G_k$ and a prime p such that $p \nmid g$, $p \notin Q$ and $p \le N^{1/z^2}$. Moreover, let

$$G_{k+1} = \mathcal{A}_{k+1} \cup \mathcal{V}_{k+1}. \tag{4.14}$$

By the primitivity of $\mathcal{A}$, for all $r_1 \le k \le z$ we have

$$\mathcal{A}_k \cap \mathcal{V}_k = \varnothing. \tag{4.15}$$

Clearly, if $r_1 \le k \le z$ and $g \in G_k$, then we have $\Omega(g) = k$, and if $q \in Q$, then $q \nmid g$. It follows from these facts that if $v \in \mathcal{V}_z \subset G_z$ and $v \le N$, then we have $v \in \mathcal{E}$. However, $\mathcal{V}_z$ also contains numbers v greater than N. Such a v is the product of an $a \in \mathcal{A}$ and at most $z - r_1 \le z$ primes, each of them $\le N^{1/z^2}$. Thus we have

$$v \le N \cdot (N^{1/z^2})^z = N^{1+1/z}. \tag{4.16}$$

It follows that

$$\mathcal{V}_z \subset \mathcal{E} \cup \{n : N < n \le N^{1+1/z}\}$$

and whence

$$S(\mathcal{E}) \ge S(\mathcal{V}_z) - \sum_{N<n\le N^{1+1/z}} \frac{1}{n}. \tag{4.17}$$

Here we have

$$\sum_{N<n\le N^{1+1/z}} \frac{1}{n} = O\left(\frac{\log N}{z}\right) = o\left(\frac{\log N}{(\log\log N)^{1/2}}\right). \tag{4.18}$$

It remains to give a lower bound for $S(\mathcal{V}_z)$.

If $r_1 < k \le z$ and $v \in \mathcal{V}_k$, then v is of the form $v = pg$ with $g \in G_{k-1}$ and a prime p such that

$$p \nmid g, p \notin Q, p \le N^{1/z^2}, \tag{4.19}$$

and for all $g \in G_{k-1}$ and p satisfying (4.19), we have $v = pg \in \mathcal{V}_k$. Since here $\Omega(v) = k$, thus v has at most k representations in this form. It follows that

$$S(\mathcal{V}_k) = \sum_{v\in\mathcal{V}_k} \frac{1}{v} \ge \frac{1}{k} \sum_{g\in G_{k-1}} \sum_{\substack{p\nmid g, p\notin Q\\ p\le N^{1/z^2}}} \frac{1}{pg} = \frac{1}{k} \sum_{g\in G_{k-1}} \frac{1}{g} \sum_{\substack{p\nmid g, p\notin Q\\ p\le N^{1/z^2}}} \frac{1}{p}. \tag{4.20}$$

By (2.2), (4.16) and Lemmas 1 and 6, here the innermost sum is

$$\sum_{\substack{p\nmid g, p\notin Q\\ p\le N^{1/z^2}}} \frac{1}{p} \ge \sum_{p\le N^{1/z^2}} \frac{1}{p} - \sum_{p\mid g} \frac{1}{p} - \sum_{q\in Q} \frac{1}{q}$$

$$= \log\log N - 2\log z + O(1) - c_6 \log\log\log N^{1+1/z} + O(1) > z - c_{11}\log z. \tag{4.21}$$

It follows from (4.14), (4.15), (4.20) and (4.21) that

$$S(\mathcal{V}_k) > \frac{z - c_{11}\log z}{k} S(\mathcal{G}_{k-1}) = \frac{z - c_{11}\log z}{k}\big(S(\mathcal{A}_{k-1}) + S(\mathcal{V}_{k-1})\big) \tag{4.22}$$

for $r_1 < k \le z$. For N large enough clearly we have

$$\frac{z - c_{11}\log z}{k} \ge \begin{cases} 1 & \text{if } k < z - c_{11}\log z \\ 1 - c_{11}\frac{\log z}{z} & \text{for all } k \le z. \end{cases} \tag{4.23}$$

It follows from (4.22) and (4.23) by a simple induction argument that

$$S(\mathcal{V}_z) > \left(1 - c_{11}\frac{\log z}{z}\right)^{c_{11}\log z} \sum_{r_1\le k<z} S(\mathcal{A}_k) = \big(1+o(1)\big)\Sigma_3. \tag{4.24}$$

(4.8) follows from (4.17), (4.18) and (4.24) and this completes the proof of Theorem 1.

5 Proof of Theorem 2

$\mathcal{P} = \{p_1, p_2, p_3, \dots\} = \{2, 3, 5, \dots\}$ denotes the set of primes and p_k is the k–th prime. For given N and k consider the following sets:

$$T_N(p_1, \dots, p_k) = \{m \in \mathbb{N}_N^* : m \mid p_1 \cdot \ldots \cdot p_k\} \text{ and}$$

$$R_N(p_1, \dots, p_k) = \{m \in \mathbb{N}_N^* : (m, p_1 \cdot \ldots \cdot p_k) = 1\}.$$

It is well–known and easy to check that

$$\lim_{N\to\infty} \frac{|R_N(p_1, \dots, p_k)|}{|\mathbb{N}_N^*|} = \prod_{i=1}^{k}\left(1 - \frac{1}{p_i+1}\right). \tag{5.1}$$

More generally, for $0 \le \alpha < \beta$ we have

$$\lim_{N\to\infty} \frac{|\alpha N \le m \le \beta N : m \in \mathbb{N}^*, (m, p_1 \cdot \ldots \cdot p_k) = 1|}{|\alpha N \le n \le \beta N : n \in \mathbb{N}^*|} = \prod_{i=1}^{k}\left(1 - \frac{1}{p_i+1}\right). \tag{5.2}$$

Clearly, every $a \in \mathbb{N}_N^*$ can be uniquely represented as $a = a_1 \cdot a_2$, where

$$a_1 \in T_N(p_1, \ldots, p_k),\ a_2 \in R_N(p_1, \ldots, p_k).$$

For any primitive set $\mathcal{A} \in \mathbb{P}_N^*$ and $a_2 \in R_N(p_1, \ldots, p_k)$ we set

$$F(a_2) = \big\{a \in \mathcal{A} : a = a_1 \cdot a_2,\ a_1 \in T_N(p_1, \ldots, p_k)\big\}.$$

Of course some of the $F(a_2)$ can be empty.

Then

$$\mathcal{A} = \dot{\bigcup_{a_2 \in R_N(p_1,\ldots,p_k)}} F(a_2),\ |\mathcal{A}| = \sum_{a_2 \in R_N(p_1,\ldots,p_k)} |F(a_2)|. \tag{5.3}$$

For $n \in \mathbb{N}$, let $\mathcal{L}_k(n)$ be a primitive set of $T_n(p_1, \ldots, p_k)$ with maximal cardinality. We observe that in the case $n < p_{k+1}$ the set $\mathcal{L}_k(n)$ is an optimal primitive squarefree set in $\mathbb{N}_n^*$, i.e.

$$\max_{\mathcal{B} \in \mathbb{P}_n^*} |\mathcal{B}| = |\mathcal{L}_k(n)|.$$

Since $\mathcal{A}$ is a primitive set then necessarily we have

$$|F(a_2)| \leq \left|\mathcal{L}_k\left(\left\lfloor \frac{N}{a_2} \right\rfloor\right)\right|$$

for every $a_2 \in R_N(p_1, \ldots, p_k)$.

Therefore by (5.3), for every primitive set $\mathcal{A} \in \mathbb{P}_N^*$

$$|\mathcal{A}| \leq \sum_{a_2 \in R_N(p_1,\ldots,p_k)} \left|\mathcal{L}_k\left(\left\lfloor \frac{N}{a_2} \right\rfloor\right)\right| \quad \text{holds.} \tag{5.4}$$

Finding in general the value $|\mathcal{L}_k(n)|$ seems difficult. However, for small k and arbitrary n this can easily be done. Take $k = 3$. We have

$$|\mathcal{L}_3(1)| = |\mathcal{L}_3(2)| = 1$$
$$|\mathcal{L}_3(3)| = |\mathcal{L}_3(4)| = 2,\ \big(\mathcal{L}_3(3) = \mathcal{L}_3(4) = \{2, 3\}\big)$$
$$|\mathcal{L}_3(n)| = 3 \text{ for } n \geq 5,\ \big(\mathcal{L}_3(n) = \{2, 3, 5\} \text{ or } \{6.10, 15\}\big).$$

Consequently, for $a_2 \in R_N(2, 3, 5)$ we obtain

$$\mathcal{L}_3\left(\left\lfloor \frac{N}{a_2} \right\rfloor\right) = 1, \text{ if } \frac{N}{3} < a_2 \leq N,$$
$$\mathcal{L}_3\left(\left\lfloor \frac{N}{a_2} \right\rfloor\right) = 2, \text{ if } \frac{N}{5} < a_2 \leq \frac{N}{3}, \text{ and}$$
$$\mathcal{L}_3\left(\left\lfloor \frac{N}{a_2} \right\rfloor\right) = 3, \text{ if } a_2 \leq \frac{N}{5}.$$

Hence by (5.4) we have

$$
\begin{aligned}
|\mathcal{A}| \leq & \left| \frac{N}{3} < a_2 \leq N : (a_2, 30) = 1, a_2 \in \mathbb{N}^* \right| \cdot 1 + \\
& \left| \frac{N}{5} < a_2 \leq \frac{N}{3} : (a_2, 30) = 1, a_2 \in \mathbb{N}^* \right| \cdot 2 + \\
& \left| a_2 \leq \frac{N}{5} : (a_2, 30) = 1, a_2 \in \mathbb{N}^* \right| \cdot 3.
\end{aligned}
$$

Using (5.2) we get

$$
\begin{aligned}
|\mathcal{A}| &\leq \left(\frac{2}{3} \cdot 1 + \frac{2}{15} \cdot 2 + \frac{1}{5} \cdot 3 \right) \cdot |\mathbb{N}_N^*| \cdot \left(1 - \frac{1}{3}\right) \left(1 - \frac{1}{4}\right) \left(1 - \frac{1}{6}\right) + o(N) \\
&= \frac{23}{36} |\mathbb{N}_N^*| + o(N) \sim 0,6389 \cdot |\mathbb{N}_N^*| + o(N).
\end{aligned}
$$

Taking $k = 5$, by similar calculations (details are omitted) we obtain a slightly better bound:

$$|\mathcal{A}| \leq 0,6366 \cdot |\mathbb{N}_N^*| + o(N).$$

This proves the upper bound (ii).

Proof of (i): Consider the following set

$$\mathcal{A} = \dot{\bigcup_{a_2 \in R_N(2,3,5)}} G(a_2),$$

where for every $a_2 \in R_N(2, 3, 5)$ $G(a_2)$ is defined by

$$
G(a_2) = \begin{cases}
\{a_2\}, & \text{if } \frac{N}{3} < a_2 \leq N \\
\{2a_2, 3a_2\}, & \text{if } \frac{N}{5} < a_2 \leq \frac{N}{3} \\
\{2a_2, 3a_2, 5a_2\}, & \text{if } \frac{N}{21} < a_2 \leq \frac{N}{5} \\
\{6a_2, 10a_2, 15a_2\}, & \text{if } \frac{N}{7 \cdot 21} < a_2 \leq \frac{N}{21} \\
\{30a_2\}, & \text{if } \frac{N}{7^2 \cdot 21} < a_2 \leq \frac{N}{7 \cdot 21} \\
\varnothing, & \text{if } a_2 \leq \frac{N}{7^2 \cdot 21}
\end{cases}
$$

It can easily be verified that $\mathcal{A} \in \mathbb{P}_N^*$.

Simple calculations, using (5.2), yield

$$|\mathcal{A}| \geq 0,6328 \cdot |\mathbb{N}_N^*| + o(N).$$

By considering the set $R_N(2, 3, 5, 7, 11)$, similarly one can construct a primitive set $\mathcal{A}'$ with a slightly better bound:

$$|\mathcal{A}'| \geq 0,6362 \cdot |\mathbb{N}_N^*| + o(N).$$

References

[1] R. Ahlswede, L.H. Khachatrian and A. Sárközy, On prefix–free and suffix–free sequences of integers, Preprint 99–050, SFB 343 “Diskrete Strukturen in der Mathematik”, Universit”at Bielefeld, Number, Information and Complexity, Special volume in honour of R. Ahlswede on occasion of his 60th birthday, editors I. Althöfer, N. Cai, G. Dueck, L. Khachatrian, M. Pinsker, A. Sárközy, I. Wegener, and Z. Zhang, Kluwer Acad. Publ. ...

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

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

[4] P. Erdős, On the integers having exactly k prime factors, Ann. Math. 49, 53–66, 1948.

[5] P. Erdős and A. Sárközy, On the number of prime factors of integers, Acta Sci. Math. (Szeged) 42, 237–246, 1980.

[6] P. Erdős, A. Sárközy and E. Szemegédi, On an extremal problem concerning primitive sequences, J. London Math. Soc. 42, 484–488, 1967.

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

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

[9] K. Prachar, Primzahlverteilung, Springer–Verlag, Berlin–Göttingen–Heidelberg, 1957.

[10] C. Pomerance and A. Sárközy, On homogeneous multiplicative hybrid problems in number theory, Acta Arith. 49, 291–302, 1988.

[11] L.G. Sathe, On a problem of Hardy on the distribution of integers having a given number of prime factors, J. Indian Math. Soc. 17, 63–141, 1953 and 18, 27–81, 1954.

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

[13] Zhenxiang Zhang, On a conjecture of Erdős on the sum $\sum_{p \le n} 1/(p \log p)$, J. Number Theory 39, 14–17, 1991.