THE ERDŐS CONJECTURE FOR PRIMITIVE SETS

JARED DUKER LICHTMAN AND CARL POMERANCE

Abstract. A subset of the integers larger than 1 is *primitive* if no member divides another. Erdős proved in 1935 that the sum of $1/(a \log a)$ for a running over a primitive set A is universally bounded over all choices for A. In 1988 he asked if this universal bound is attained for the set of prime numbers. In this paper we make some progress on several fronts, and show a connection to certain prime number "races" such as the race between $\pi(x)$ and $\mathrm{li}(x)$.

1. Introduction

A set of positive integers > 1 is called primitive if no element of A divides any other (for convenience, we exclude the singleton set $\{1\}$). There are a number of interesting and sometimes unexpected theorems about primitive sets. After Besicovitch [2], we know that the upper asymptotic density of a primitive set can be arbitrarily close to $1/2$, whereas the lower asymptotic density is always 0. Using the fact that if a primitive set has a finite reciprocal sum, then the set of multiples of members of the set has an asymptotic density, Erdős gave an elementary proof that the set of nondeficient numbers (i.e., $\sigma(n)/n \geq 2$, where σ is the sum-of-divisors function) has an asymptotic density. Though the reciprocal sum of a primitive set can possibly diverge, Erdős [7] showed that for a primitive set A,

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

In fact, the proof shows that these sums are uniformly bounded as A varies over primitive sets.

Some years later in a 1988 seminar in Limoges, Erdős suggested that in fact we always have

$$f(A) := \sum_{a \in A} \frac{1}{a \log a} \leq \sum_{p \in \mathcal{P}} \frac{1}{p \log p}, \tag{1.1}$$

where $\mathcal{P}$ is the set of prime numbers. The assertion (1.1) is now known as the Erdős conjecture for primitive sets.

In 1991, Zhang [18] proved the Erdős conjecture for primitive sets A with no member having more than 4 prime factors (counted with multiplicity).

After Cohen [5], we have

$$C := \sum_{p \in \mathcal{P}} \frac{1}{p \log p} = 1.63661632336\ldots, \tag{1.2}$$

Date: June 6, 2018.
2010 *Mathematics Subject Classification.* Primary 11B83; Secondary 11A05, 11N05.
Key words and phrases. primitive set, primitive sequence, Mertens' product formula.

the sum over primes in (1.1). Using the original Erdős argument in [7], Erdős and Zhang showed that $f(A) < 2.886$ for a primitive set A, which was later improved by Robin to 2.77. These unpublished estimates are reported in Erdős–Zhang [9] who used another method to show that $f(A) < 1.84$. Shortly after, Clark [4] claimed that $f(A) \le e^\gamma = 1.781072\ldots$. However, his brief argument appears to be incomplete.

Our principal results are the following.

Theorem 1.1. *For any primitive set A we have $f(A) < e^\gamma$.*

Theorem 1.2. *For any primitive set A with no element divisible by 8, we have $f(A) < C + 10^{-5}$.*

Say a prime p is **Erdős strong** if for any primitive set A with the property that each element of A has least prime factor p, we have $f(A) \le 1/(p\log p)$. We conjecture that every prime is Erdős strong. Note that the Erdős conjecture (1.1) would immediately follow, though it is not clear that the Erdős conjecture implies our conjecture. Just proving our conjecture for the case of $p = 2$ would go a long way toward solving the problem. Currently the best we can do for a primitive set A of even numbers is that $f(A) < e^\gamma/2$, see Proposition 2.1 below.

For part of the next result, we assume the Riemann hypothesis (RH) and the Linear Independence hypothesis (LI), which asserts that the sequence of numbers $\gamma_n > 0$ such that $\zeta(\frac{1}{2} + i\gamma_n) = 0$ is linearly independent over $\mathbb{Q}$.

Theorem 1.3. *Unconditionally, all of the odd primes among the first 10^8 primes are Erdős strong. Assuming RH and LI, the Erdős strong primes have relative lower logarithmic density > 0.995.*

The proof depends strongly on a recent result of Lamzouri [12] who was interested in the "Mertens race" between $\prod_{p\le x}(1 - 1/p)$ and $1/(e^\gamma \log x)$.

For a primitive set A, let $\mathcal{P}(A)$ denote the support of A, i.e., the set of prime numbers that divide some member of A. It is clear that the Erdős conjecture (1.1) is equivalent to the same assertion where the prime sum is over $\mathcal{P}(A)$.

Theorem 1.4. *If A is a primitive set with $\mathcal{P}(A) \subset [3, \exp(10^6)]$, then*

$$f(A) \le \sum_{p\in\mathcal{P}(A)} \frac{1}{p\log p}.$$

If some primitive set A of odd numbers exists with $f(A) > \sum_{p\in\mathcal{P}(A)} 1/(p\log p)$, Theorem 1.4 suggests that it will be very difficult indeed to give a concrete example!

For a positive integer n, let $\Omega(n)$ denote the number of prime factors of n counted with multiplicity. Let $\mathbb{N}_k$ denote the set of integers n with $\Omega(n) = k$. Zhang [19] proved a result that implies $f(\mathbb{N}_k) < f(\mathbb{N}_1)$ for each $k \ge 2$, so that the Erdős conjecture holds for the primitive sets $\mathbb{N}_k$. More recently, Banks and Martin [1] conjectured that $f(\mathbb{N}_1) > f(\mathbb{N}_2) > f(N_3) > \cdots$. The inequality $f(\mathbb{N}_2) > f(\mathbb{N}_3)$ was just established by Bayless, Kinlaw, and Klyve [3]. We prove the following result.

Theorem 1.5. *There is a positive constant c such that $f(\mathbb{N}_k) \ge c$ for all k.*

We let the letters p, q, r represent primes. In addition, we let p_n represent the nth prime. For an integer $a > 1$, we let $P(a)$ and $p(a)$ denote the largest and smallest

prime factors of a. Modifying the notation introduced in [9], for a primitive set A let

$$A_p = \{a \in A : p(a) \geq p\},$$
$$A'_p = \{a \in A : p(a) = p\},$$
$$A''_p = \{a/p : a \in A'_p\}.$$

We let $f(a) = 1/(a \log a)$ and so $f(A) = \sum_{a \in A} f(a)$. In this language, Zhang's full result [19] states that $f((\mathbb{N}_k)'_p) \leq f(p)$ for all primes p, $k \geq 1$. We also, let

$$g(a) = \frac{1}{a} \prod_{p < P(a)} \left(1 - \frac{1}{p}\right), \quad h(a) = \frac{1}{a \log P(a)},$$

with $g(A) = \sum_{a \in A} g(a)$ and $h(A) = \sum_{a \in A} h(a)$.

2. The Erdős approach

In this section we will prove Theorem 1.1. We begin with an argument inspired by the original 1935 paper of Erdős [7].

Proposition 2.1. *For any primitive set A, if $q \notin A$ then*

$$f(A'_q) < e^{\gamma} g(q) = \frac{e^{\gamma}}{q} \prod_{p < q} \left(1 - \frac{1}{p}\right).$$

Proof. For each $a \in A'_q$, let $S_a = \{ba : p(b) \geq P(a)\}$. Note that S_a has asymptotic density $g(a)$. Since A'_q is primitive, we see that the sets S_a are pairwise disjoint. Further, the union of the sets S_a is contained in the set of all natural numbers m with $p(m) = q$, which has asymptotic density $g(q)$. Thus, the sum of densities for each S_a is dominated by $g(q)$, that is,

$$g(A'_q) = \sum_{a \in A'_q} g(a) \leq g(q). \tag{2.1}$$

By Theorem 7 in [14], we have for $x \geq 285$,

$$\prod_{p \leq x} \left(1 - \frac{1}{p}\right) > \frac{1}{e^{\gamma} \log(2x)}, \tag{2.2}$$

which may be extended to all $x \geq 1$ by a calculation. Thus, since each $a \in A'_q$ is composite,

$$g(a) = \frac{1}{a} \prod_{p < P(a)} \left(1 - \frac{1}{p}\right) > \frac{e^{-\gamma}}{a \log\big(2P(a)\big)} > \frac{e^{-\gamma}}{a \log a} = e^{-\gamma} f(a).$$

Hence by (2.1),

$$f(A'_q)/e^{\gamma} < g(A'_q) \leq g(q). \tag{2.3}$$

□

Remark 2.2. Let σ denote the sum-of-divisors function and let A be the set of n with $\sigma(n)/n \geq 2$ and $\sigma(d)/d < 2$ for all proper divisors d of n, the set of primitive nondeficient numbers. Then $g(A)$ gives the density of nondeficient numbers, recently shown in [11] to lie in the tight interval (0.2476171, 0.2476475). In [13], an analog of Proposition 2.1 is a key ingredient for sharp bounds on the reciprocal sum of the primitive nondeficient numbers.

Remark 2.3. We have $g(\mathcal{P}) = 1$. It is easy to see by induction over primes r that

$$\sum_{p\le r} g(p) = \sum_{p\le r} \frac{1}{p} \prod_{q<p} \left(1 - \frac{1}{q}\right) = 1 - \prod_{p\le r} \left(1 - \frac{1}{p}\right).$$

Letting $r \to \infty$ we get that $g(\mathcal{P}) = 1$. As a consequence, we have

$$\sum_{p>2} g(p) = \frac{1}{2}, \tag{2.4}$$

an identity we will find to be useful.

For a primitive set A, let

$$A^k = \{a : 2^k \| a \in A\}, \qquad B^k = \{a/2^k : a \in A^k\}.$$

The next result will help us prove Theorem 1.1.

Lemma 2.4. *For a primitive set A, let $k \ge 1$ be such that $2^k \notin A$. Then we have*

$$f(A^k) < \frac{e^\gamma}{2^k} \sum_{\substack{p\notin A\\ p>2}} g(p).$$

Proof. If $p2^k \notin A$ for a prime $p > 2$, then $(B^k)'_p$ is a primitive set of odd composite numbers, so by Proposition 2.1, $f((B^k)'_p) < e^\gamma g(p)$.

Now if $p2^k \in A$ for some odd prime p, then $(B^k)'_p = \{p\}$ and note $p \notin A$ by primitivity. We have $f(p2^k) < 2^{-k} e^\gamma g(p)$ since

$$\frac{1}{p2^k \log(p2^k)} \le \frac{1}{p2^k \log(2p)} < \frac{e^\gamma}{2^k} g(p),$$

which follows from (2.2). Hence combining the two cases,

$$f(A^k) = \sum_{\substack{p\notin A\\ p>2}} f(2^k \cdot (B^k)'_p) \le \sum_{\substack{p\in B^k, p\notin A\\ p>2}} f(p2^k) + 2^{-k} \sum_{\substack{p\notin B^k\\ p>2}} f((B^k)'_p) < \frac{e^\gamma}{2^k} \sum_{\substack{p\notin A\\ p>2}} g(p).$$

□

With Lemma 2.4 in hand, we prove $f(A) < e^\gamma$.

Proof of Theorem 1.1. From Erdős–Zhang [9], we have that $f(A_3) < 0.92$. If $2 \in A$, then $A'_2 = \{2\}$, so that $f(A) = f(A_3) + f(A'_2) < 0.92 + 1/(2\log 2) < e^\gamma$. Hence we may assume that $2 \notin A$. If A contains every odd prime, then $f(A'_2)$ consists of at most one power of 2, and the calculation just concluded shows we may assume this is not the case. Hence there is at least one odd prime $p \notin A$. By Proposition 2.1, we have

$$f(A) = \sum_p f(A'_p) = \sum_{p\in A} f(p) + \sum_{p\notin A} f(A'_p) < \sum_{p\in A} f(p) + e^\gamma \sum_{\substack{p\notin A\\ p>2}} g(p) + f(A'_2). \tag{2.5}$$

First suppose A contains no powers of 2. Then by Lemma 2.4,

$$f(A'_2) = \sum_{k\ge 1} f(A^k) < \sum_{k\ge 1} \frac{e^\gamma}{2^k} \sum_{\substack{p\notin A\\ p>2}} g(p) = e^\gamma \sum_{\substack{p\notin A\\ p>2}} g(p).$$

Substituting into (2.5), we conclude, using (2.4),

$$f(A) < \sum_{p\in A} f(p) + 2e^{\gamma} \sum_{\substack{p\notin A\\ p>2}} g(p) \le 2e^{\gamma} \sum_{p>2} g(p) = e^{\gamma}. \tag{2.6}$$

For the last inequality we used that for every prime p,

$$\frac{f(p)}{e^{\gamma} g(p)} < 1.082, \tag{2.7}$$

which follows after a short calculation using [14, Theorem 7].

Now if $2^K \in A$ for some positive integer K, then K is unique and $K \ge 2$. Also $A^K = \{2^K\}$ and $A^k = \emptyset$ for all $k > K$, so again by Lemma 2.4,

$$f(A_2') = \sum_{k=1}^{K} f(A^k) = f(2^K) + \sum_{k=1}^{K-1} \frac{e^{\gamma}}{2^k} \sum_{\substack{p\notin A\\ p>2}} g(p) = f(2^K) + (1-2^{1-K})e^{\gamma} \sum_{\substack{p\notin A\\ p>2}} g(p).$$

Substituting into (2.5) gives

$$\begin{aligned} f(A) &< \sum_{p\in A} f(p) + f(2^K) + (2-2^{1-K})e^{\gamma} \sum_{\substack{p\notin A\\ p>2}} g(p) \le f(2^K) + (2-2^{-1})e^{\gamma} \sum_{p>2} g(p) \\ &\le f(2^2) + (1-2^{-2})e^{\gamma} < e^{\gamma}, \end{aligned} \tag{2.8}$$

using $K \ge 2$, the identity (2.4), inequality (2.7), and $f(2^2) < 2^{-2}e^{\gamma}$. This completes the proof. □

3. Mertens primes

In this section we will prove Theorems 1.3 and Theorem 1.4. Note that by Mertens' theorem,

$$\prod_{p<x}\left(1-\frac{1}{p}\right) \sim \frac{1}{e^{\gamma}\log x}, \quad x\to\infty,$$

where γ is Euler's constant. We say a prime q is **Mertens** if

$$e^{\gamma}\prod_{p<q}\left(1-\frac{1}{p}\right) \le \frac{1}{\log q}, \tag{3.1}$$

and let $\mathcal{P}^{\mathrm{Mert}}$ denote the set of Mertens primes. We are interested in Mertens primes because of the following consequence of Proposition 2.1, which shows that every Mertens prime is Erdős strong.

Corollary 3.0.1. *Let A be a primitive set. If $q \in \mathcal{P}^{\mathrm{Mert}}$, then $f(A_q') \le f(q)$. Hence if $A_q' \subset \{q\}$ for all $q \notin \mathcal{P}^{\mathrm{Mert}}$, then A satisfies the Erdős conjecture.*

Proof. By (2.7) we have $f(A_q') \le \max\{e^{\gamma}g(q), f(q)\}$. If $q \in \mathcal{P}^{\mathrm{Mert}}$, then

$$e^{\gamma}g(q) = \frac{e^{\gamma}}{q}\prod_{p<q}\left(1-\frac{1}{p}\right) \le \frac{1}{q\log q} = f(q),$$

so $f(A_q') \le f(q)$. □

Now, one would hope that the Mertens inequality (3.1) holds for all primes q. However, (3.1) fails for $q = 2$ since $e^{\gamma} > 1/\log 2$. We have computed that q is indeed a Mertens prime for all $2 < q \le p_{10^8} = 2{,}038{,}074{,}743$, thus proving the unconditional part of Theorem 1.3.

3.1. **Proof of Theorem 1.3.** To complete the proof, we use a result of Lamzouri [12] relating the Mertens inequality to the race between $\pi(x)$ and $\mathrm{li}(x)$, studied by Rubinstein and Sarnak [17]. Under the assumption of RH and LI, he proved that the set $\mathcal{N}$ of real numbers x satisfying

$$e^{\gamma}\prod_{p\le x}\left(1-\frac{1}{p}\right) > \frac{1}{\log x},$$

has logarithmic density $\delta(\mathcal{N})$ equal to the logarithmic density of numbers x with $\pi(x) > \mathrm{li}(x)$, and in particular

$$\delta(\mathcal{N}) = \lim_{x\to\infty}\frac{1}{\log x}\int_{t\in\mathcal{N}\cap[2,x]}\frac{dt}{t} = 0.00000026\ldots. \tag{3.2}$$

We note that if a prime $p = p_n \in \mathcal{N}$, then for $p' = p_{n+1}$ we have $[p, p') \subset \mathcal{N}$ because the prime product on the left-hand side is constant on $[p, p')$, while $1/\log x$ is decreasing for $x \in [p, p')$.

The set of primes $\mathcal{Q}$ in $\mathcal{N}$ is precisely the set of non-Mertens primes, so $\mathcal{Q} = \mathcal{P} \setminus \mathcal{P}^{\mathrm{Mert}}$. From the above observation, we may leverage knowledge of the continuous logarithmic density $\delta(\mathcal{N})$ to obtain an upper bound on the relative (upper) logarithmic density of non-Mertens primes

$$\bar{\delta}(\mathcal{Q}) := \limsup_{x\to\infty}\frac{1}{\log x}\sum_{\substack{p\le x\\ p\in\mathcal{Q}}}\frac{\log p}{p}. \tag{3.3}$$

From the above observation, we have

$$\delta(\mathcal{N}) \ge \limsup_{x\to\infty}\frac{1}{\log x}\sum_{\substack{p\le x\\ p\in\mathcal{Q}}}\int_p^{p'}\frac{dt}{t} = \limsup_{x\to\infty}\frac{1}{\log x}\sum_{\substack{p\le x\\ p\in\mathcal{Q}}}\log(p'/p).$$

Then letting $d_p = p' - p$ be the gap between consecutive primes, we have

$$\delta(\mathcal{N}) \ge \limsup_{x\to\infty}\frac{1}{\log x}\sum_{\substack{p\le x\\ p\in\mathcal{Q}}}\frac{d_p}{p},$$

since $\sum \log(p'/p) = \sum d_p/p + O(1)$. The average gap is roughly $\log p$, so we may consider the primes for which $d_p < \epsilon \log p$, for a small positive constant ϵ to be determined.

We claim

$$\limsup_{x\to\infty}\frac{1}{\log x}\sum_{\substack{p\le x\\ d_p<\epsilon\log p}}\frac{\log p}{p} \;\le\; 16\epsilon, \tag{3.4}$$

from which it follows

$$\bar{\delta}(\mathcal{Q}) = \limsup_{x\to\infty}\frac{1}{\log x}\sum_{\substack{p\le x\\ p\in\mathcal{Q}}}\frac{\log p}{p} \le \limsup_{x\to\infty}\frac{1}{\log x}\Big(\sum_{\substack{p\le x\\ p\in\mathcal{Q}\\ d_p\ge\epsilon\log p}}\frac{d_p/\epsilon}{p} + \sum_{\substack{p\le x\\ d_p<\epsilon\log p}}\frac{\log p}{p}\Big)$$
$$\le \delta(\mathcal{N})/\epsilon + 16\epsilon.$$

Hence to prove Theorem 1.3 it suffices to prove (3.4), since taking $\epsilon = \sqrt{\delta(\mathcal{N})}/4$ gives

$$\bar{\delta}(\mathcal{Q}) < 8\sqrt{\delta(\mathcal{N})} < 4.2 \times 10^{-3}. \tag{3.5}$$

By Riesel-Vaughan [16, Lemma 5], the number of primes p up to x with $p+d$ also prime is at most

$$\sum_{\substack{p \le x \\ p+d \text{ prime}}} 1 \le \frac{8c_2 x}{\log^2 x} \prod_{\substack{p \mid d \\ p>2}} \frac{p-1}{p-2},$$

where c_2 is for the twin-prime constant $2\prod_{p>2} p(p-2)/(p-1)^2 = 1.3203\ldots$. Denote the prime product by $F(d) = \prod_{\substack{p\mid d \\ p>2}} \frac{p-1}{p-2}$, and consider the multiplicative function $H(d) = \sum_{u\mid d} \mu(u)F(d/u)$. We have $H(2^k) = 0$ for all $k \ge 1$, and for $p > 2$ we have $H(p) = F(p) - 1$, and $H(p^k) = 0$ if $k \ge 2$. Thus,

$$\begin{aligned}\sum_{d\le y} F(d) &= \sum_{d\le y}\sum_{u\mid d} H(u) = \sum_{u\le y} H(u) \sum_{d\le y/u} 1 \le y \sum_{u\le y} \frac{H(u)}{u} \le y \prod_{p>2}\Big(1 + \frac{H(p)}{p}\Big) \\ &= y\prod_{p>2}\Big(1 + \frac{(p-1)/(p-2)-1}{p}\Big) = y\prod_{p>2}\Big(1+\frac{1}{p(p-2)}\Big).\end{aligned}$$

Noting that $c_2' := \prod_{p>2}(1 + 1/[p(p-2)]) = 2/c_2$, we have

$$\sum_{\substack{p\le x \\ d_p < \epsilon \log p}} 1 \le \sum_{d \le \epsilon \log x} \sum_{\substack{p \le x \\ p+d \text{ prime}}} 1 \le \frac{8c_2 x}{\log^2 x} \sum_{d\le \epsilon \log x} F(d) \le \epsilon \frac{8c_2 c_2' x}{\log x} = \epsilon \frac{16x}{\log x}.$$

Thus, (3.4) now follows by partial summation, and the proof is complete.

Remark 3.1. The concept of relative upper logarithmic density of the set of non-Mertens primes in (3.3) can be replaced in the theorem with

$$\bar{\delta}_0(\mathcal{Q}) := \limsup_{x\to\infty} \frac{1}{\log\log x} \sum_{\substack{p\le x \\ p\in\mathcal{Q}}} \frac{1}{p}.$$

Indeed, $\bar{\delta}_0(\mathcal{Q}) \le \bar{\delta}(\mathcal{Q})$ follows from the identity

$$\sum_{\substack{p\le x \\ p\in\mathcal{Q}}} \frac{1}{p} = \frac{1}{\log x} \sum_{\substack{p\le x \\ p\in\mathcal{Q}}} \frac{\log p}{p} + \int_2^x \frac{1}{t(\log t)^2} \sum_{\substack{p\le t \\ p\in\mathcal{Q}}} \frac{\log p}{p}\, dt.$$

Remark 3.2. Greg Martin has indicated to us that one should be able to prove (under RH and LI) that the relative logarithmic density of $\mathcal{Q}$ exists and is equal to the logarithmic density of $\mathcal{N}$. The idea is as follows. Partition the positive reals into intervals of the form $[y, y+y^{1/3})$. Let E_1 be the union of those intervals $[y, y+y^{1/3})$ where the sign of $e^\gamma \prod_{p\le x}(1-1/p) - 1/\log x$ is not constant and let E_2 be the union of those intervals $[y, y+y^{1/3})$ which do not have $\sim y^{1/3}/\log y$ primes as $y \to \infty$. The the logarithmic density of $E_1 \cup E_2$ can be shown to be 0, from which the assertion follows.

3.2. **Proof of Theorem 1.4.** We now use some numerical estimates of Dusart [6] to prove Theorem 1.4.

We say a pair of primes $p \le q$ is a **Mertens pair** if

$$\prod_{p \le r < q} \left(1 - \frac{1}{r}\right) > \frac{\log p}{\log pq}.$$

We claim that every pair of primes p, q with $2 < p \le q < e^{10^6}$ is a Mertens pair. Assume this and let A be a primitive set supported on the odd primes to e^{10^6}. By (2.1), if $p \notin A$, we have

$$\begin{aligned} \frac{1}{p} &\ge \sum_{a \in A'_p} \frac{1}{a} \prod_{p \le r < P(a)} \left(1 - \frac{1}{r}\right) > \sum_{a \in A'_p} \frac{\log p}{a \log(p\, P(a))} \\ &\ge \sum_{a \in A'_p} \frac{\log p}{a \log a} = f(A'_p) \log p. \end{aligned}$$

Dividing by $\log p$ we obtain $f(A'_p) \le f(p)$, which also holds if $p \in A$. Thus, the claim about Mertens pairs implies the theorem.

To prove the claim, first note that if p is a Mertens prime, then p, q is a Mertens pair for all primes $q \ge p$. Indeed, we have

$$\prod_{p \le r < q} \left(1 - \frac{1}{r}\right) = \prod_{r < p} \left(1 - \frac{1}{r}\right)^{-1} \prod_{r < q} \left(1 - \frac{1}{r}\right) > e^{\gamma} \log p \prod_{r < q} \left(1 - \frac{1}{r}\right).$$

By (2.2), this last product exceeds $e^{-\gamma}/\log(2q) > e^{-\gamma}/\log(pq)$, and using this in the above display shows that p, q is indeed a Mertens pair. Since all of the odd primes up to p_{10^8} are Mertens, to complete the proof of our assertion, it suffices to consider the case when $p > p_{10^8}$. Define E_p via the equation

$$\prod_{r < p} \left(1 - \frac{1}{r}\right) = \frac{1 + E_p}{e^{\gamma} \log p}.$$

Using [6, Theorem 5.9], we have $|E_p| \le .2/(\log p)^3$. A routine calculation shows that if $p \le q < e^{4.999(\log p)^4}$, then

$$\prod_{p \le r < q} \left(1 - \frac{1}{r}\right) = \frac{\log p}{\log q} \cdot \frac{1 + E_q}{1 + E_p} > \frac{\log p}{\log pq}.$$

It remains to note that $4.999(\log p_{10^8})^4 > 1{,}055{,}356$.

It seems interesting to record the principle that we used in the proof.

Corollary 3.2.1. *If A is a primitive set such that $p(a), P(a)$ is a Mertens pair for each $a \in A$, then $f(A) \le f(\mathcal{P}(A))$.*

Remark 3.3. Kevin Ford has noted to us the remarkable similarity between the concept of Mertens primes in this paper and the numbers

$$\gamma_n = \left(\gamma + \sum_{k \le n} \frac{\log p_k}{p_k - 1}\right) \prod_{k \le n} \left(1 - \frac{1}{p_k}\right)$$

discussed in Diamond–Ford [10]. In particular, while it may not be obvious from the definition, the analysis in [10] on whether the sequence $\gamma_1, \gamma_2, \ldots$ is monotone is quite similar to the analysis in [12] on the Mertens inequality. Though the numerical

evidence seems to indicate we always have $\gamma_{n+1} < \gamma_n$, this is disproved in [10], and it is indicated there that the first time this fails may be near $1.9 \cdot 10^{215}$. This may also be near where the first odd non-Mertens prime exists. If this is the case, and under assumption of RH, it may be that every pair of primes $p \le q$ is a Mertens pair when $p > 2$ and $q < \exp(10^{100})$.

4. Odd primitive sets

In this section we prove a somewhat stronger version of Theorem 1.2.

Let

$$K(y) = y + \log y + \frac{\log y - 1.1}{y}, \qquad L(y) = y + \log y - 1 + \frac{\log y - 2.1}{y}.$$

By Proposition 5.16 in Dusart [6], we have

$$\begin{aligned} p_n &> n(\log n + \log\log n - 1 + (\log\log n - 2.1)/\log n = nL(\log n) && n \ge 2, \\ \log p_n &> n + \log L(\log n) > K(\log n) && n \ge 315. \end{aligned}$$

It follows that

$$p_n \log p_n > nL(\log n)K(\log n), \quad n \ge 315. \tag{4.1}$$

We have the following result.

Lemma 4.1. *For any primitive set A, we have $f(A_{p_n}) < 1/K(\log n)$ for all $n \ge e^{12}$.*

Proof. We first prove an inequality similar to [9, Lemma 2]:

$$\sum_{m \ge n} \frac{1}{mL(\log m)K(\log m)} < \frac{1}{K(\log n)}, \quad n \ge e^{12}. \tag{4.2}$$

Indeed, we have

$$\begin{aligned} \sum_{m \ge n} \frac{1}{mL(\log m)K(\log m)} &< \frac{1}{nL(\log n)K(\log n)} + \int_n^\infty \frac{dt}{tL(\log t)K(\log t)} \\ &= \frac{1}{nL(\log n)K(\log n)} + \int_{\log n}^\infty \frac{dy}{L(y)K(y)}. \end{aligned}$$

We will show that

$$\frac{1}{nL(\log n)K(\log n)} < \int_{\log n}^\infty \frac{dy}{y^2 L(y)K(y)^2}, \quad n \ge 120 \tag{4.3}$$

and that

$$\frac{1}{L(y)K(y)} + \frac{1}{y^2 L(y)K(y)^2} < \frac{K'(y)}{K(y)^2}, \quad y \ge 11.957. \tag{4.4}$$

Assuming these estimates, (4.2) follows, since

$$\sum_{m \ge n} \frac{1}{mL(\log m)K(\log m)} < \int_{\log n}^\infty \frac{K'(y)}{K(y)^2}\, dy = \frac{1}{K(\log n)}.$$

To show (4.3), one may check it numerically for $n \in [120, 2000]$, after which one has $L(y), K(y) \in [y, 2y]$ for all $y \ge \log 2000$, and thus

$$\int_{\log n}^\infty \frac{dy}{y^2 L(y)K(y)^2} > \int_{\log n}^\infty \frac{dy}{8y^5} = \frac{(\log n)^{-4}}{32} > \frac{(\log n)^{-2}}{n} > \frac{1}{nL(\log n)K(\log n)}.$$

Additionally, (4.4) is equivalent to

$$\begin{aligned} 0 &< K'(y)\cdot L(y) - K(y) - 1/y^2 \\ &= \frac{0.1}{y} - \frac{5.2}{y^2} - \frac{(y+1)\log^2 y - (4.1y+4.2)\log y + 4.41}{y^3}, \end{aligned}$$

which is seen to hold for $y \geq 11.957$.

For a primitive set A, the **degree** of A is $d^\circ(A) = \max\{\Omega(a) : a \in A\}$. As in [9], we may assume that $d^\circ(A)$ is finite. With the key inequality (4.2) in hand, we prove the lemma by induction on $k = d^\circ(A_{p_n})$. If $d^\circ(A_{p_n}) \leq 1$, then by (4.1) and (4.2),

$$f(A_{p_n}) \leq \sum_{m\geq n} f(p_m) < \sum_{m\geq n} \frac{1}{mL(\log m)K(\log m)} < \frac{1}{K(\log n)}. \tag{4.5}$$

Now let $k \geq 2$, assume $d^\circ(A_{p_m}) = k$, and suppose the lemma holds for all cases with degree $< k$. We have $f(A_{p_n}) = \sum_{m\geq n} f(A'_{p_m})$. If $d^\circ(A'_{p_m}) \leq 1$ then

$$f(A'_{p_m}) \leq \frac{1}{p_m \log p_m} < \frac{1}{p_m K(\log m)}.$$

If $d^\circ(A'_{p_m}) > 1$, then $d^\circ(A''_{p_m}) \in [1, k-1]$ so that by the induction hypothesis, $f(A''_{p_m}) < 1/K(\log m)$. But $f(A'_{p_m}) < f(A''_{p_m})/p_m$, so that in either case, we have

$$f(A'_{p_m}) < \frac{1}{p_m K(\log m)} < \frac{1}{mL(\log m)K(\log m)}.$$

The result then follows from (4.2). □

Combining with the fact that $f(A'_{p_n}) \leq 1/(p_n \log p_n)$ for $n \in [2, 10^8]$, since such p_n are Mertens, we have that

$$f(A_3) < \frac{1}{K(\log 10^8)} + \sum_{2\leq n\leq 10^8} \frac{1}{p_n \log p_n}.$$

Thus, since one may compute $\sum_{m\leq 10^8} 1/(p_m \log p_m) = 1.589964286878896\ldots$, we have proved the following result.

Theorem 4.2. *For any odd primitive set $A = A_3$, we have*

$$f(A) < \epsilon_0 + \sum_{p\geq 3} \frac{1}{p\log p} \tag{4.6}$$

where

$$\epsilon_0 := \frac{1}{K(\log 10^8)} - C + \sum_{n\leq 10^8} \frac{1}{p_n \log p_n} \leq 5.86\times 10^{-6}.$$

This theorem yields the following corollary.

Corollary 4.2.1. *If A is a primitive set containing no multiple of 8, then*

$$f(A) < \epsilon_0 + \sum_{p\in\mathcal{P}(A)} \frac{1}{p\log p}.$$

Proof. First suppose that all members of A are odd. Let Q be the set of odd primes not in $\mathcal{P}(A)$. Applying Theorem 4.2 to $A \cup Q$ we obtain the corollary in this case. Next, suppose that A contains an even number, but no multiple of 4. If $2 \in A$, the result follows by applying the corollary to $A \setminus \{2\}$, so assume $2 \notin A$. Then A_2'' is an odd primitive set and $f(A_2') < f(A_2'')/2$. Let $\mathcal{P}' = \mathcal{P}(A) \setminus \{2\}$. We have by the odd case that

$$f(A) = f(A_3) + f(A_2') < \epsilon_0 + \frac{3}{2} \sum_{p \in \mathcal{P}'} \frac{1}{p \log p} < \sum_{p \in \mathcal{P}(A)} \frac{1}{p \log p},$$

which is stronger than required. The case when A contains a multiple of 4 but no multiple of 8 follows in a similar fashion, but with $\frac{3}{2}$ above replaced with $\frac{7}{4}$. □

We remark that $f(A) \leq C + \epsilon_0$ for all primitive sets A of cube-free numbers.

5. Zhang primes and the Banks–Martin conjecture

Note that

$$\sum_{p \geq x} \frac{1}{p \log p} \sim \frac{1}{\log x}, \quad x \to \infty.$$

In Erdős–Zhang [9] and in Zhang [19], numerical approximations to this asymptotic relation are used, in a similar manner as (4.5) above. Say a prime q is **Zhang** if

$$\sum_{p \geq q} \frac{1}{p \log p} \leq \frac{1}{\log q}.$$

Let $\mathcal{P}^{\mathrm{Zh}}$ denote the set of Zhang primes. We are interested in Zhang primes because of the following result.

Theorem 5.1. *If $\mathcal{P}(A_p') \subset \mathcal{P}^{Zh}$, then $f(A_p') \leq f(p)$. Hence the Erdős conjecture holds for all primitive sets A supported on $\mathcal{P}^{Zh}$.*

Proof. Recall the notation d° defined in the proof of Lemma 4.1. We proceed by induction on $d^\circ(A_p')$. If $d^\circ(A_p') \leq 1$, then $f(A_p') \leq f(p)$. If $d^\circ(A_p') > 1$, then $f(A_p') \leq f(A_p'')/p$. The primitive set $B := A_p''$ satisfies $f(B) = f(B_p) = \sum_{q \geq p} f(B_q')$. Since $d^\circ(B_q') \leq d^\circ(B) < d^\circ(A_p')$, by induction we have $f(B_q') \leq f(q)$. Thus, since p is Zhang,

$$f(A_p'') = f(B) = \sum_{q \geq p} f(B_q') \leq \sum_{q \geq p} \frac{1}{q \log q} \leq \frac{1}{\log p},$$

from which we obtain $f(A_p') \leq f(A_p'')/p \leq 1/(p \log p)$. This completes the proof. □

From this one might hope that all primes are Zhang. However, the prime 2 is not Zhang since $C > 1/\log 2$, and the prime 3 is not Zhang since $C - 1/(2 \log 2) > 1/\log 3$. Nevertheless, as with Mertens primes, it is true that the remaining primes up to p_{10^8} are Zhang. Indeed, starting from (1.2), we computed that

$$\sum_{p \geq q} \frac{1}{p \log p} = C - \sum_{p < q} \frac{1}{p \log p} \leq \frac{1}{\log q} \qquad \text{for all } 3 < q \leq p_{10^8}. \tag{5.1}$$

The computation stopped at 10^8 for convenience, and one could likely extend this further with some patience. It seems likely that there is also a "race" between $\sum_{p \geq q} 1/(p \log p)$ and $1/\log q$, as with Mertens primes, and that a large logarithmic

density of primes q are Zhang, with a small logarithmic density of primes failing to be Zhang.

A related conjecture due to Banks and Martin [1] is the chain of inequalities,

$$\sum_p \frac{1}{p\log p} > \sum_{p\le q} \frac{1}{pq\log pq} > \sum_{p\le q\le r} \frac{1}{pqr\log pqr} > \cdots,$$

succinctly written as $f(\mathbb{N}_k) > f(\mathbb{N}_{k+1})$ for all $k \ge 1$, where $\mathbb{N}_k = \{n : \Omega(n) = k\}$. As mentioned in the introduction, we know only that $f(\mathbb{N}_1) > f(\mathbb{N}_k)$ for all $k \ge 2$ and $f(\mathbb{N}_2) > f(\mathbb{N}_3)$. More generally, for a subset Q of primes, let $\mathbb{N}_k(Q)$ denote the subset of $\mathbb{N}_k$ supported on Q. A result of Zhang [19] impies that $f(\mathbb{N}_1(Q)) > f(\mathbb{N}_k(Q))$ for all $k > 1$, while Banks and Martin showed that $f(\mathbb{N}_k(Q)) > f(\mathbb{N}_{k+1}(Q))$ if $\sum_{p\in Q} 1/p$ is not too large. We prove a similar result in the case where Q is a subset of the Zhang primes and we replace $f(\mathbb{N}_k(Q))$ with $h(\mathbb{N}_k(Q))$. Recall $h(A) = \sum_{a\in A} 1/(a\log P(a))$.

Proposition 5.2. *For all $k \ge 1$ and $Q \subset \mathcal{P}^{Zh}$, we have $h(\mathbb{N}_k(Q)) \ge h(\mathbb{N}_{k+1}(Q))$.*

Proof. Since p_k is a Zhang prime, we have

$$\begin{aligned} h(\mathbb{N}_{k+1}(Q)) &= \sum_{\substack{q_1\le\cdots\le q_{k+1}\\ q_i\in Q}} \frac{1}{q_1\cdots q_k q_{k+1}\log q_{k+1}} \\ &= \sum_{\substack{q_1\le\cdots\le q_k\\ q_i\in Q}} \frac{1}{q_1\cdots q_k} \sum_{q_{k+1}\ge q_k} \frac{1}{q_{k+1}\log q_{k+1}} \\ &\le \sum_{\substack{q_1\le\cdots\le q_k\\ q_i\in Q}} \frac{1}{q_1\cdots q_k\log q_k} = h(\mathbb{N}_k(Q)). \end{aligned}$$

This completes the proof. □

It is interesting that if we do not in some way restrict the primes used, the analogue of the Banks–Martin conjecture for the function h fails. In particular, we have

$$h(\mathbb{N}_2) > \sum_{m\le 10^4} \frac{1}{p_m} \sum_{n\ge m} \frac{1}{p_n\log p_n} = \sum_{m\le 10^4} \frac{1}{p_m}\left(C - \sum_{k<m} \frac{1}{p_k\log p_k}\right) > 1.638,$$

while $h(\mathbb{N}_1) = C < 1.637$.

It is also interesting that the analogue of the Banks–Martin conjecture for the function g is false since

$$1 = g(\mathbb{N}_1) = g(\mathbb{N}_2) = g(\mathbb{N}_3) = \cdots.$$

We have already shown in (2.1) that $g(A'_q) \le g(q)$ for any primitive set A and prime q, so the analogue for g of the strong Erdős conjecture holds.

5.1. **Proof of Theorem 1.5.** We now return to the function f and prove Theorem 1.5.

We may assume that k is large. Let $m = \lfloor \sqrt{k} \rfloor$ and let $B(n) = e^{e^n}$. We have

$$f(\mathbb{N}_k) = \sum_{\Omega(a)=k} \frac{1}{a \log a} > \sum_{\substack{\Omega(a)=k \\ e^{e^k} < a \le e^{e^{k+m}}}} \frac{1}{a \log a}$$

$$= \sum_{j \le m} \sum_{\substack{\Omega(a)=k \\ B(k+j-1) < a \le B(k+j)}} \frac{1}{a \log a} > \sum_{j \le m} \frac{1}{\log B(k+j)} \sum_{\substack{\Omega(a)=k \\ B(k+j-1) < a \le B(k+j)}} \frac{1}{a}.$$

Thus it suffices to show that there is a positive constant c such that for $j \le m$ we have

$$\sum_{\substack{\Omega(a)=k \\ B(k+j-1) < a \le B(k+j)}} \frac{1}{a} \ge c \frac{\log B(k+j)}{m} = c \frac{e^{k+j}}{m}, \tag{5.2}$$

so that the proposition will follow.

Let $N_k(x)$ denote the number of members of $\mathbb{N}_k$ in $[1, x]$. We use the Sathe–Selberg theorem, see [15, Theorem 7.19], from which we have that uniformly for $B(k) < x \le B(k+m)$, as $k \to \infty$,

$$N_k(x) \sim \frac{x}{k!} \frac{(\log \log x)^k}{\log x}.$$

This result also follows from Erdős [8].

We have

$$\sum_{\substack{\Omega(a)=k \\ B(k+j-1) < a \le B(k+j)}} \frac{1}{a} > \int_{B(k+j-1)}^{B(k+j)} \frac{N_k(x) - N_k(B(k+j-1))}{x^2} \, dx$$

$$\gg \int_{2B(k+j-1)}^{B(k+j)} \frac{N_k(x)}{x^2} \, dx.$$

Thus,

$$\sum_{\substack{\Omega(a)=k \\ B(k+j-1) < a \le B(k+j)}} \frac{1}{a} \gg \frac{(\log \log B(k+j-1))^k}{k!} \int_{2B(k+j-1)}^{B(k+j)} \frac{dx}{x \log x}$$

$$= \frac{(k+j-1)^k}{k!} (\log \log B(k+j) - \log \log(2B(k+j-1))$$

$$\gg \frac{(k+j-1)^k}{k!} \gg \frac{e^{k+j}}{\sqrt{k}},$$

the last estimate following from Stirling's formula. This proves (5.2), and so the theorem.

The sets $\mathbb{N}_k$ and Theorem 1.5 give us the following result.

Corollary 5.2.1. *We have that*

$$\limsup_{x \to \infty} \{ f(A) : A \subset [x, \infty),\ A \text{ primitive} \} > 0.$$

Acknowledgments

We thank Greg Martin for the content of Remark 3.2 and Kevin Ford for the content of Remark 3.3. We thank Paul Kinlaw and Zhenxiang Zhang for some helpful comments.

References

1. W. D. Banks and G. Martin, *Optimal primitive sets with restricted primes*, Integers **13** (2013), #A69, 10 pp.
2. A. S. Besicovitch, *On the density of certain sequences of integers*, Math. Ann. **110** (1934), 336–341.
3. J. Bayless, P. Kinlaw, and D. Klyve, *Sums over primitive sets with a fixed number of prime factors*, preprint, 2018.
4. D. A. Clark, *An upper bound of* $\sum 1/(a_i \log a_i)$ *for primitive sequences*, Proc. Amer. Math. Soc. **123** (1995), 363–365.
5. H. Cohen, *High precision computation of Hardy-Littlewood constants*, preprint `https://www.math.u-bordeaux.fr/~hecohen/` .
6. P. Dusart, *Explicit estimates of some functions over primes*, Ramanujan J. **45** (2018), 227–251.
7. P. Erdős, *Note on sequences of integers no one of which is divisible by any other*, J. London Math. Soc. **10** (1935), 126–128.
8. ———, *On the integers having exactly k prime factors*, Annals Math. **49** (1948), 53–66.
9. P. Erdős and Z. Zhang, *Upper bound of* $\sum 1/(a_i \log a_i)$ *for primitive sequences*, Proc. Amer. Math. Soc. **117** (1993), 891–895.
10. H. G. Diamond and K. Ford, *Generalized Euler constants*, Math. Proc. Cambridge Phil. Soc. **145** (2008), 27–41.
11. M. Kobayashi, *On the density of abundant numbers*, PhD thesis, Dartmouth College, 2010.
12. Y. Lamzouri, *A bias in Mertens' product formula*, Int. J. Number Theory **12** (2016), 97–109.
13. J. D. Lichtman, *The reciprocal sum of primitive nondeficient numbers*, J. Number Theory (2018), `https://doi.org/10.1016/j.jnt.2018.03.021`.
14. J. B. Rosser and L. Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois J. Math. **6** (1962), 64–94.
15. H. L. Montgomery and R. C. Vaughan, *Multiplicative number theory I. Classical theory*, Cambridge U. Press, Cambridge, 2007.
16. H. Riesel and R. C. Vaughan, *On sums of primes*, Ark. Mat. **21** (1983), 45–74.
17. M. Rubinstein and P. Sarnak, *Chebyshev's bias*, Experiment. Math. **3** (1994), 173–197.
18. Z. Zhang, *On a conjecture of Erdős on the sum* $\sum_{p \le n} 1/(p \log p)$, J. Number Theory **39** (1991), 14–17.
19. ———, *On a problem of Erdős concerning primitive sequences*, Math. Comp. **60** (1993), 827–834.

Department of Mathematics, Dartmouth College, Hanover, NH 03755
E-mail address: `jdl.18@dartmouth.edu`
E-mail address: `jared.d.lichtman@gmail.com`

Department of Mathematics, Dartmouth College, Hanover, NH 03755
E-mail address: `carl.pomerance@dartmouth.edu`