PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY, SERIES B
Volume 6, Pages 1–14 (April 10, 2019)
https://doi.org/10.1090/bproc/40

THE ERDŐS CONJECTURE FOR PRIMITIVE SETS

JARED DUKER LICHTMAN AND CARL POMERANCE

(Communicated by Amanda Folsom)

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 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 [4], 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., $\sigma(n)/n \ge 2$, where σ is the sum-of-divisors function) has an asymptotic density. Though the reciprocal sum of a primitive set can possibly diverge, Erdős [9] 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

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

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

Received by the editors June 6, 2018, and, in revised form, June 26, 2018, August 3, 2018, and August 14, 2018.

2010 *Mathematics Subject Classification.* Primary 11B83; Secondary 11A05, 11N05.

Key words and phrases. Primitive set, primitive sequence, Mertens' product formula.

The first-named author is grateful for support from the office of undergraduate research at Dartmouth College.

After Cohen [6], we have

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

the sum over primes in (1.1). Using the original Erdős argument in [9], Erdős and Zhang showed that $f(A) < 2.89$ for a primitive set A, which was later improved by Robin to 2.77. These unpublished estimates are reported in [11] by Erdős–Zhang, who used another method to show that $f(A) < 1.84$. Shortly after, Clark [5] 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 $f(A) < C + 2.37 \times 10^{-7}$.*

Say a prime p is *Erdős strong* if for any primitive set A with the property that each element of A has the same 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 give the inequality in Theorem 1.2 for all primitive sets A. 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 [13], 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 [20] 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 [2] conjectured that $f(\mathbb{N}_1) > f(\mathbb{N}_2) > f(\mathbb{N}_3) > \cdots$. 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 $f(\mathbb{N}_k) \geq c$ 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 [11], for a primitive set A let

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

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 [20] states that $f((\mathbb{N}_k)'_p) \leq f(p)$ for all primes p, $k \geq 1$. We also let

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

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 [9].

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 [17], 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,

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

Hence by (2.1),

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

∎

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 an appropriate analog of $g(A)$ gives the density of nondeficient numbers recently shown in [12] to lie in the tight interval $(0.2476171, 0.2476475)$. In [14], 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 $g(\mathcal{P}) = 1$. Indeed, it is easy to see by induction over primes r that

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

Letting $r\to\infty$ we get that $g(\mathcal{P}) = 1$. There is also a holistic way of seeing this. Since $g(p)$ is the density of the set of integers with least prime factor p, it would make sense that $g(\mathcal{P})$ is the density of the set of integers which have a least prime factor, and this density is 1. To make this rigorous, one notes that the density of the set of integers whose least prime factor is $> y$ tends to 0 as $y\to\infty$. As a consequence of $g(\mathcal{P}) = 1$, we have

$$\sum_{p>2} g(p) = \frac{1}{2}, \tag{2.3}$$

an identity we will find to be useful.

For a primitive set A, let

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

The next result will help us prove Theorem 1.1.

Lemma 2.4. *For a primitive set A, let $k\geq 1$ 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. The hypothesis $2^k\notin A$ implies that $1\notin B^k$, so that B^k is a primitive set. If $2^k p\notin A$ 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 $2^k p\in A$ for some odd prime p, then $(B^k)'_p = \{p\}$, and note that $p\notin A$ by primitivity. We have $f(2^k p) < 2^{-k} e^\gamma g(p)$ since

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

which follows from (2.2). Since $(B^k)'_p \neq \emptyset$ implies $p\notin A$, we have

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

$\square$

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

Proof of Theorem 1.1. From Erdős–Zhang [11], 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 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 $p_0 \notin A$. By Proposition 2.1, we have

$$f(A) = \sum_p f(A'_p) = \sum_{p\in A} f(p) + \sum_{p\notin A} f(A'_p) < \sum_{p\in A} f(p) + e^\gamma \sum_{\substack{p\notin A\\ p>2}} g(p) + f(A'_2). \tag{2.4}$$

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

$$f(A'_2) = \sum_{k\ge 1} f(A^k) < \sum_{k\ge1} \frac{e^\gamma}{2^k} \sum_{\substack{p\notin A\\ p>2}} g(p) = e^\gamma \sum_{\substack{p\notin A\\ p>2}} g(p).$$

Substituting into (2.4), we conclude, using (2.3), that

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

For the last inequality in (2.5) we used that for every prime p,

$$\frac{f(p)}{e^\gamma g(p)} < 1.082, \tag{2.6}$$

which follows after a short calculation using [17, 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 (2.4) gives

$$\begin{aligned} f(A) &< \sum_{p\in A} f(p) + f(2^K) + (2-2^{1-K})e^\gamma \sum_{\substack{p\notin A\\ p>2}} g(p) \le f(2^K) + (2-2^{1-K})e^\gamma \sum_{p>2} g(p) \\ &\le f(2^K) + (1-2^{-K})e^\gamma < e^\gamma, \end{aligned} \tag{2.7}$$

using identity (2.3), inequality (2.6), and $f(2^K) < 2^{-K}e^\gamma$. This completes the proof. □

3. Mertens primes

In this section we will prove Theorems 1.3 and 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 $\mathcal{P}^{\rm Mert}$ 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 3.1. *Let A be a primitive set. If $q \in \mathcal{P}^{\rm Mert}$, then $f(A'_q) \le f(q)$. Hence if $A'_q \subset \{q\}$ for all $q \notin \mathcal{P}^{\rm Mert}$, then A satisfies the Erdős conjecture.*

Proof. By Proposition 2.1 we have $f(A'_q) \le \max\{e^\gamma g(q), f(q)\}$. If $q \in \mathcal{P}^{\rm Mert}$, 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 $e^\gamma > 1/\log 2$. Nevertheless, 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 [13] relating the Mertens inequality to the race between $\pi(x)$ and $\mathrm{li}(x)$, studied by Rubinstein and Sarnak [18]. 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 $\mathcal{Q} = \mathcal{P}\setminus\mathcal{P}^{\rm Mert}$. 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 that

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

from which it follows that

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

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

$$\sum_{\substack{p\le x\\ p+d \text{ prime}}} 1 \le \frac{8c_2 x}{\log^2 x} \prod_{\substack{p\mid d\\ p>2}} \frac{p-1}{p-2}, \tag{3.6}$$

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,

$$\sum_{d\le y} F(d) = \sum_{d\le y}\sum_{u\mid d} H(u) = \sum_{u\le y} H(u) \sum_{d\le y/u} 1 \le y\sum_{u\le y} \frac{H(u)}{u} \le y\prod_{p>2}\Big(1+\frac{H(p)}{p}\Big)$$
$$= y\prod_{p>2}\Big(1+\frac{(p-1)/(p-2)-1}{p}\Big) = y\prod_{p>2}\Big(1+\frac{1}{p(p-2)}\Big) = \frac{2y}{c_2}.$$

Using (3.6), we have

$$\sum_{\substack{p\le x\\ d_p<\epsilon\log p}} 1 \le \sum_{d\le\epsilon\log x} \sum_{\substack{p\le x\\ p+d \text{ prime}}} 1 \le \frac{8c_2 x}{\log^2 x} \sum_{d\le\epsilon\log x} F(d) \le \epsilon\frac{8c_2c_2' x}{\log x} = \epsilon\frac{16x}{\log x}.$$

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

Remark 3.2. 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 3.3. Greg Martin has indicated to us that one may 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 $\mathcal{N}$. This topic will be addressed in a future paper.

3.2. **Proof of Theorem 1.4.** We now use some numerical estimates of Dusart [8] 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 up 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 [8, Theorem 5.9], we have for $p > 2{,}278{,}382$,

$$|E_p| \le .2/(\log p)^3. \tag{3.7}$$

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 3.4. *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 3.5. 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 [7]. In particular, while it may not be obvious from the definition, the analysis in [7] on whether the sequence $\gamma_1, \gamma_2, \ldots$ is monotone is quite similar to the analysis in [13] on the Mertens inequality. Though the numerical evidence seems to indicate we always have $\gamma_{n+1} < \gamma_n$, this is disproved in [7], 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 $q < \exp(3 \cdot 10^{11})$.

4. Odd primitive sets

We say a primitive set is odd if every member of the set is an odd number. In this section we prove Theorem 1.2 and establish a curious result on parity for primitive sets.

Let

$$\epsilon_0 = \sum_{\substack{p>2\\ p\notin\mathcal{P}^{\mathrm{Mert}}}} \left(e^\gamma g(p) - f(p)\right).$$

Lemma 4.1. *We have $0 \le \epsilon_0 < 2.37 \times 10^{-7}$.*

Proof. By the definition of $\mathcal{P}^{\mathrm{Mert}}$, the summands in the definition of ϵ_0 are non-negative, so that $\epsilon_0 \ge 0$. If $p > 2$ is not Mertens, then $p > p_{10^8} > 2 \times 10^9$, so that (3.7) shows that

$$e^\gamma g(p) - f(p) < \frac{1}{5p(\log p)^4}. \tag{4.1}$$

By [8, Proposition 5.16], we have

$$p_n > n(\log n + \log\log n - 1 + (\log\log n - 2.1)/\log n, \quad n \ge 2.$$

Using this we find that

$$\sum_{n>10^8} \frac{1}{5p_n(\log p_n)^4} < 2.37 \times 10^{-7},$$

which with (4.1) completes the proof. □

Remark 4.2. Clearly, a smaller bound for ϵ_0 would follow by raising the search limit for Mertens primes. Another small improvement could be made using the estimate in [1] for p_n. It follows from the ideas in Remark 3.3 that $\epsilon_0 > 0$. Further, it may be provable from the ideas in Remark 3.5 that $\epsilon_0 < 10^{-12}$ if the Riemann Hypothesis holds.

We have the following result.

Theorem 4.3. *For any odd primitive set A, we have*

$$f(A) \le f(\mathcal{P}(A)) + \epsilon_0. \tag{4.2}$$

Proof. Assume that A is an odd primitive set. If $p \in \mathcal{P}(A)$ is Mertens, Corollary 3.1 implies that $f(A'_p) \le f(p)$, while if $p \in \mathcal{P}(A)$ is not Mertens we have by Proposition 2.1 that $f(A'_p) \le \max\{f(p), e^\gamma g(p)\} = e^\gamma g(p)$. Thus,

$$f(A) = \sum_{p\in\mathcal{P}(A)} f(A'_p) \le \sum_{p\in\mathcal{P}(A)\cap\mathcal{P}^{\mathrm{Mert}}} f(p) + \sum_{p\in\mathcal{P}(A)\setminus\mathcal{P}^{\mathrm{Mert}}} e^\gamma g(p) \le \epsilon_0 + \sum_{p\in\mathcal{P}(A)} f(p)$$

by the definition of ϵ_0. This completes the proof. □

This theorem yields the following corollary.

Corollary 4.4. *If A is a primitive set containing no multiple of* 8*, then* (4.2) *holds.*

Proof. We have seen the corollary in the case that A is odd. Next, suppose that A contains an even number but no multiple of 4. If $2 \in A$, the result follows by applying Theorem 4.3 to $A \setminus \{2\}$, so assume $2 \notin A$. Then A''_2 is an odd primitive set and $f(A'_2) \le f(A''_2)/2$. We have by the odd case that

$$f(A) = f(A_3) + f(A'_2) < f(\mathcal{P}(A_3)) + \epsilon_0 + \frac{1}{2}\left(f(\mathcal{P}(A''_2)) + \epsilon_0\right). \tag{4.3}$$

Since

$$\frac{1}{2}f(\mathcal{P}(A''_2)) \le \frac{1}{2}f(\mathcal{P}\setminus\{2\}) < 0.4577$$

and $f(2) = 0.7213\ldots$, (4.3) and Lemma 4.1 imply that $f(A) < f(\mathcal{P}(A))$, which is stronger than required. The case when A contains a multiple of 4 but no multiple of 8 follows in a similar fashion. □

Since a cube-free number cannot be divisible by 8, (4.2) holds for all primitive sets A of cube-free numbers. Also, the proof of Corollary 4.4 can be adapted to show that (4.2) holds for all primitive sets A containing no number that is 4 (mod 8).

We close out this section with a curious result about those primitive sets A where (4.2) does not hold. Namely, the Erdős conjecture must then hold for the set of odd members of A. Put another way, (4.2) holds for any primitive set A for which the Erdős conjecture for the odd members of A *fails*.

Theorem 4.5. *If A is a primitive set with $f(A) > f(\mathcal{P}(A)) + \epsilon_0$, then $f(A_3) < f(\mathcal{P}(A_3))$.*

Sketch of proof. Without loss of generality, we may include in A all primes not in $\mathcal{P}(A)$ and so assume that $\mathcal{P}(A) = \mathcal{P}$ and $f(A) > C + \epsilon_0$. By Theorem 4.3 we may assume that A is not odd, and by Corollary 4.4 we may assume that $2 \notin A$. By the proof of Theorem 1.1 (see (2.5) and (2.7)), if $3 \in A$, we have

$$f(A) < f(3) + \frac{2}{3}e^\gamma < C,$$

a contradiction, so we may assume that $3 \notin A$. We now apply the method of proof of Theorem 1.1 to A_3, where powers of 3 replace powers of 2. This leads to

$$f(A_3) < \frac{1}{2}e^\gamma < C - f(2) = f(\mathcal{P}(A_3)).$$

This completes the argument. □

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 [11] and in Zhang [20], numerical approximations to this asymptotic relation are exploited. 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. As in [11] it suffices to prove the theorem in the case that A is a finite set. By $d^\circ(A)$ we mean the maximal value of $\Omega(a)$ for $a\in A$. 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 [2] is the chain of inequalities
$$\sum_{p}\frac{1}{p\log p}>\sum_{p\leq q}\frac{1}{pq\log pq}>\sum_{p\leq q\leq r}\frac{1}{pqr\log pqr}>\cdots,$$
succinctly written as $f(\mathbb{N}_k)>f(\mathbb{N}_{k+1})$ for all $k\geq 1$, 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 $k\geq 2$ 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 [20] implies 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 that $h(A)=\sum_{a\in A}1/(a\log P(a))$.

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 the primes in Q are Zhang primes, 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 analog 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 analog 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 analog 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

$$\begin{aligned} f(\mathbb{N}_k) &= \sum_{\Omega(a)=k} \frac{1}{a \log a} > \sum_{\substack{\Omega(a)=k \\ e^{e^k} < a \le e^{e^{k+m}}}} \frac{1}{a \log a} \\ &= \sum_{j \le m} \sum_{\substack{\Omega(a)=k \\ B(k+j-1) < a \le B(k+j)}} \frac{1}{a \log a} > \sum_{j \le m} \frac{1}{\log B(k+j)} \sum_{\substack{\Omega(a)=k \\ B(k+j-1) < a \le B(k+j)}} \frac{1}{a}. \end{aligned}$$

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

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

since 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 [10].

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,

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

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 5.3. *We have that*

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

Acknowledgments

We thank Greg Martin for his thoughts in connection with Remark 3.3 and Kevin Ford for the content of Remark 3.5. We thank Paul Kinlaw, Zhenxiang Zhang, and the referee for some helpful comments.

References

[1] C. Axler, *New estimates for the n-th prime number*, arXiv:1706.03651v1 [math.NT], 2017.

[2] William D. Banks and Greg Martin, *Optimal primitive sets with restricted primes*, Integers **13** (2013), Paper No. A69, 10. MR3118387

[3] J. Bayless, P. Kinlaw, and D. Klyve, *Sums over primitive sets with a fixed number of prime factors*, Math. Comp., electronically published on March 5, 2019, DOI:10.1090/mcom/3416 (to appear in print).

[4] A. S. Besicovitch, *On the density of certain sequences of integers*, Math. Ann. **110** (1935), no. 1, 336–341, DOI 10.1007/BF01448032. MR1512943

[5] David A. Clark, *An upper bound of $\sum 1/(a_i \log a_i)$ for primitive sequences*, Proc. Amer. Math. Soc. **123** (1995), no. 2, 363–365, DOI 10.2307/2160889. MR1243164

[6] H. Cohen, *High precision computation of Hardy-Littlewood constants*, preprint, `https://www.math.u-bordeaux.fr/~hecohen/` .

[7] Harold G. Diamond and Kevin Ford, *Generalized Euler constants*, Math. Proc. Cambridge Philos. Soc. **145** (2008), no. 1, 27–41, DOI 10.1017/S0305004108001187. MR2431637

[8] Pierre Dusart, *Explicit estimates of some functions over primes*, Ramanujan J. **45** (2018), no. 1, 227–251, DOI 10.1007/s11139-016-9839-4. MR3745073

[9] Paul Erdös, *Note on sequences of integers no one of which is divisible by any other*, J. London Math. Soc. **10** (1935), no. 2, 126–128, DOI 10.1112/jlms/s1-10.1.126. MR1574239

[10] P. Erdös, *On the integers having exactly K prime factors*, Ann. of Math. (2) **49** (1948), 53–66, DOI 10.2307/1969113. MR0023279

[11] Paul Erdős and Zhen Xiang Zhang, *Upper bound of* $\sum 1/(a_i \log a_i)$ *for primitive sequences*, Proc. Amer. Math. Soc. **117** (1993), no. 4, 891–895, DOI 10.2307/2159512. MR1116257

[12] Mitsuo Kobayashi, *On the density of abundant numbers*, ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–Dartmouth College, 2010. MR2996025

[13] Youness Lamzouri, *A bias in Mertens' product formula*, Int. J. Number Theory **12** (2016), no. 1, 97–109, DOI 10.1142/S1793042116500068. MR3455269

[14] Jared Duker Lichtman, *The reciprocal sum of primitive nondeficient numbers*, J. Number Theory **191** (2018), 104–118, DOI 10.1016/j.jnt.2018.03.021. MR3825463

[15] Hugh L. Montgomery and Robert C. Vaughan, *Multiplicative number theory. I. Classical theory*, Cambridge Studies in Advanced Mathematics, vol. 97, Cambridge University Press, Cambridge, 2007. MR2378655

[16] H. Riesel and R. C. Vaughan, *On sums of primes*, Ark. Mat. **21** (1983), no. 1, 46–74, DOI 10.1007/BF02384300. MR706639

[17] J. Barkley Rosser and Lowell Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois J. Math. **6** (1962), 64–94. MR0137689

[18] Michael Rubinstein and Peter Sarnak, *Chebyshev's bias*, Experiment. Math. **3** (1994), no. 3, 173–197. MR1329368

[19] Zhen Xiang Zhang, *On a conjecture of Erdős on the sum* $\sum_{p\le n} 1/(p \log p)$, J. Number Theory **39** (1991), no. 1, 14–17, DOI 10.1016/0022-314X(91)90030-F. MR1123165

[20] Zhen Xiang Zhang, *On a problem of Erdős concerning primitive sequences*, Math. Comp. **60** (1993), no. 202, 827–834, DOI 10.2307/2153122. MR1181335

Department of Mathematics, Dartmouth College, Hanover, New Hampshire 03755
Email address: jdl.18@dartmouth.edu
Email address: jared.d.lichtman@gmail.com

Department of Mathematics, Dartmouth College, Hanover, New Hampshire 03755
Email address: carl.pomerance@dartmouth.edu