A NOTE ON PROBLEM #311

QUANYU TANG

1. Introduction

In [2, p. 40], Erdős and Graham posed the following question, which also appears as Problem #311 on Bloom's Erdős Problems website [1].

Problem 1.1. *Let $\delta(N)$ be the minimal value of $|1 - \sum_{n\in A} \frac{1}{n}|$ as A ranges over all subsets of $\{1,\dots,N\}$ which contain no S such that $\sum_{n\in S}\frac{1}{n} = 1$? Is it*

$$e^{-(c+o(1))N}$$

for some constant $c \in (0,1)$?

A trivial lower bound is

$$\delta(N) \ge \frac{1}{\operatorname{lcm}(1,2,\dots,N)}.$$

In this note, we refine a representation lemma of Liu–Sawhney [3, Lemma 4.1] and obtain the following nontrivial upper bound: there exists a constant $c_0 > 0$ such that, for all sufficiently large N,

$$\delta(N) \le \exp\left(-c_0 \frac{N}{(\log N)^3(\log\log N)^3}\right).$$

2. preliminaries

2.1. **Notation.** As is standard, we write $e(x) = e(2\pi i x)$. We let $\mathrm{Re}(x)$ denote the real part of x. Given $n = p_1^{a_1}\cdots p_k^{a_k}$ where p_i are distinct primes, we let $\omega(n) = k$, $\Omega(n) = \sum_{i=1}^k a_i$ and (nonstandardly) define $\widetilde{\Omega}(n) = \max_i a_i$. We say that a positive integer n is S-smooth if every prime power $q|n$ satisfies $q \le S$. Throughout the paper, we will without comment use p to indicate a prime variable and q to denote a prime power variable.

Consider a set of integers $A \subseteq [M,N]$ where $M = N^{1-o(1)}$ and $\sum_{n\in A}\frac{1}{n} = \eta\log(N/M)$; e.g. A has logarithmic density η in the interval $[M,N]$. Let $\mathcal{Q}_A$ denote the set of prime power factors dividing at least one element of A and Q denote the least common multiple of the integers in $\mathcal{Q}_A$. Note that the sum of unit fractions with reciprocals in A may be written as a fraction with denominator Q.

Furthermore we let $R(A) = \sum_{n\in A}\frac{1}{a}$. Given a set A, we let $A_d = \{n \in A : d|n\}$. Throughout let $[a,b] = \{a,\dots,b\}$ for integers $a \le b$. For a set $\mathcal{Q}$ (normally of prime powers), we let $[\mathcal{Q}]$ denote the least common multiple.

We use standard asymptotic notation. Given functions $f = f(n)$ and $g = g(n)$, we write $f = O(g)$, $f \ll g$, or $g \gg f$ to mean that there is a constant C such that $|f(n)| \le Cg(n)$ for sufficiently large n. We write $f \asymp g$ or $f = \Theta(g)$ to mean that $f \ll g$ and $g \ll f$, and write $f = o(g)$ to mean $f(n)/g(n) \to 0$ as $n \to \infty$.

Date: January 14, 2026.

2.2. **Definition.** For a positive integer N, define

$$\delta(N) := \min\left\{ \left| 1 - \sum_{n\in A} \frac{1}{n} \right| : \ A \subseteq \{1, \dots, N\} \text{ and } A \text{ is admissible} \right\},$$

where we say that A is admissible if it contains no subset $S \subseteq A$ with $\sum_{n\in S} \frac{1}{n} = 1$. Note that any A with $\sum_{n\in A} \frac{1}{n} < 1$ is automatically admissible, since then $\sum_{n\in S} \frac{1}{n} < 1$ for every $S \subseteq A$.

2.3. **A standard input.**

Lemma 2.1. *Let*

$$\psi(x) := \sum_{p^k \le x} \log p$$

be Chebyshev's function. Then for every real $x \ge 1$,

$$\log \mathrm{lcm}(1, 2, \dots, \lfloor x \rfloor) = \psi(x). \tag{2.1}$$

Moreover, by the prime number theorem (equivalently $\psi(x) \sim x$), there exist absolute constants $x_0 \ge 2$ and $c_\psi > 0$ such that for all $x \ge x_0$,

$$\psi(x) \ge c_\psi x. \tag{2.2}$$

3. A strengthened version of Liu–Sawhney Lemma 4.1

In this section, we establish a strengthened form of [3, Lemma 4.1]. Before proving this result, we first establish a refinement of a special case of [3, Proposition 3.2], which will serve as a key preliminary ingredient in the argument.

Proposition 3.1. *There exists an absolute constant $C \ge 1$ such that the following holds. Let N be sufficiently large. Consider parameters $N^{0.9999} \le S \le K \le M \le N/10$ satisfying:*

$$S \le \min\left(\frac{M^2}{CN}, \frac{K^3}{CN^2(\log\log N)^3} \right), \ \text{and } K = 10^{-7} N (\log N)^{-1},$$

Let $A = \{n \in [M, N] : n \text{ is } S\text{−smooth}, \widetilde{\Omega}(n) \le 5 \log\log N, \Omega(N) \le 10 \log\log N\}$, and let $1/9 \le p_M, \dots, p_N \le 1/2$. Then B be sampled where $n \in A$ is contained in B with probability p_n, independently. Let $\mathcal{Q}_A = \{q \le S\}$ and $Q = [\mathcal{Q}_A]$. If $x \in [1, Q]$ is an integer such that $\mathbb{E}[R(B)] = x/Q$, then $\mathbb{P}[R(B) = x/Q] \ge 1/(4Q)$.

Remark 3.2. Before presenting the proof of this proposition, we briefly explain how it differs from [3, Proposition 3.2] and from the proof given there.

(1) In the statement of Proposition 3.1, we assume the stronger uniform bounds
$$\frac{1}{9} \le p_M, \dots, p_N \le \frac{1}{2},$$
whereas in [3, Proposition 3.2] the lower bounds on $p_M, \dots, p_N$ are $(\log\log N)^{-1}$. This strengthening is only needed because, in the application of Proposition 3.1 to the proof of Lemma 3.3, the weights p_n we use are in fact bounded below by the absolute constant $1/9$. We also let $K = 10^{-7} N (\log N)^{-1}$.

(2) In the step *"Minor arcs, upper bound"* of the proof of [3, Proposition 3.2], the authors choose
$$t = \frac{100\, N^2 \log N (\log\log N)^2}{K^2},$$
whereas in our argument we take
$$t = \frac{900\, N^2 \log N \log\log N}{K^2}.$$

This modification does not affect the validity of the minor–arc upper bound. Indeed, the proof in [3] relies on the lower bound $p_n \geq (\log\log N)^{-1}$, while in our setting we have the stronger bound $p_n \geq 1/9$, which allows us to save one factor of $\log\log N$.

(3) In the step *"Minor arcs, finishing the proof"* of [3, Proposition 3.2], an averaging argument is used to produce a prime p' satisfying

$$\gcd(p', q) = 1, \qquad p' \leq \frac{K}{q}, \qquad \text{and} \qquad \left|A_{qp'} \setminus T_q\right| \leq \frac{\log N}{1000(\log\log N)^2}.$$

In contrast, we work on a suitable dyadic interval for p', which avoids the appearance of an $\Omega(n)$ factor in the corresponding double–counting argument. This refinement allows us to reduce the admissible size of p' by a factor of $\log\log N$.

(4) In the step *"Minor arcs, finishing the proof"* of [3, Proposition 3.2], a crude bound of the form $\omega(\cdot) \leq \Omega(\cdot) \ll \log\log N$ is used to estimate how many primes p can be excluded by a single "bad" element, leading to the conclusion

$$|\mathcal{P}_q| \geq \tfrac{9}{10}|\mathcal{P}|.$$

In our setting, since $m \asymp K$ and $K \asymp N/\log N$, we have $n/m \ll \log N$. Consequently, each bad element can exclude at most one prime p. This improvement allows us to raise the relevant threshold to $\log N/(\log\log N)$, yielding an additional saving of a factor of $\log\log N$ in the final estimates.

We are now ready to present the following.

Proof of Proposition 3.1. Recall the following identity for integers $a, b \neq 0$:

$$1_{a/b\in\mathbb{Z}} = \frac{1}{b} \sum_{-b/2<h\leq b/2} e\left(\frac{ha}{b}\right).$$

We can write:

$$\begin{aligned}
\mathbb{P}[R(B) - x/Q \in \mathbb{Z}] &= \sum_{B\subseteq A} \prod_{n\in B} p_n \prod_{n\in A\setminus B} (1-p_n) \cdot \frac{1}{Q} \sum_{-Q/2<h\leq Q/2} e\left(\sum_{n\in B} \frac{h}{n} - \frac{hx}{Q}\right) \\
&= \frac{1}{Q} \sum_{-Q/2<h\leq Q/2} e\left(-\frac{hx}{Q}\right) \prod_{n\in A} (1 - p_n + p_n e(h/n)) \\
&= \frac{1}{Q} \sum_{-Q/2<h\leq Q/2} \mathrm{Re}\left(e\left(-\frac{hx}{Q}\right) \prod_{n\in A} (1 - p_n + p_n e(h/n))\right), \qquad (3.1)
\end{aligned}$$

The final line holds because the expression is a probability (or alternatively by noting that the expression conjugates upon negating h).

Our ultimate goal is to establish that (3.1) is at least $\frac{1}{2Q}$. Before proving this, let us see why the proposition then follows. It suffices to prove that $\mathbb{P}[|R(B) - x/Q| \geq 1] < \frac{1}{4Q}$. Because $\mathbb{E}[R(B)] = x/Q$, the previous probability is at most $2\exp(-M^2/(2N))$ by Azuma-Hoeffding ([3, Lemma 2.6]). By [3, Theorem 2.1], we have $Q \leq \prod_{q\leq S} \ll e^{5S}$. For our choice of S, M, it holds that $2\exp(-M^2/(2N)) \leq e^{-6S}$ for $N \gg 1$.

The remainder of the proof is devoted to estimating the expression in (3.1). Via [3, Lemma 3.1] (applying [3, Lemma 3.3] to lower bound the size of A), we find that

$$\frac{1}{Q} \sum_{|h|\leq M/2} \mathrm{Re}\left(e\left(-\frac{hx}{Q}\right) \prod_{n\in A} (1 - p_n + p_n e(h/n))\right) \geq \frac{3}{4Q}.$$

The remainder of the analysis consider h such that $|h| > M/2$; e.g. the minor arcs.

Minor arcs, upper bound: For each n, let h_n be the unique integer in $(-n/2, n/2]$ with $h \equiv h_n \bmod n$. Let

$$t = \frac{900N^2 \log N \log\log N}{K^2}.$$

Finally, let $I_h = (h - K/2, h + K/2)$ be an interval and define

$$\mathcal{D}_h = \{q \in \mathcal{Q}_A : |\{n \in A_q : h_n \geq K/2\}| < t\}.$$

We now establish a relationship between the Fourier coefficient at h and the size of $|\mathcal{Q}_A \setminus \mathcal{D}_h|$. Recall that each $n \in A$ satisfies $\Omega(n) \leq 10 \log\log N$. Therefore using the second item of [3, Fact 2.5], we have

$$\begin{aligned}
\left(\prod_{n\in A} |(1 - p_n + p_n e(h/n))|\right)^{10\log\log N} &\leq \prod_{q\in\mathcal{Q}_A} \prod_{n\in A_q} |(1 - p_n + p_n e(h/n))| \\
&\leq \prod_{q\in\mathcal{Q}_A\setminus\mathcal{D}_h} \exp\left(- p_n \cdot \frac{K^2}{N^2} \cdot t\right) \\
&\leq \exp(-100|\mathcal{Q}_A \setminus \mathcal{D}_h| \log N \log\log N)
\end{aligned}$$

because $p_n \geq 1/9$, and thus

$$\prod_{n\in A} |(1 - p_n + p_n e(h/n))| \leq N^{-10|\mathcal{Q}_A\setminus\mathcal{D}_h|}. \tag{3.2}$$

Minor arcs, establishing divisibility: The crucial step in our proof is establishing there exists an element $x \in I_h$ such that $[\mathcal{D}_h] | x$.

For $q \in \mathcal{D}_h$, define

$$T_q = \{n \in A_q : |h_n| < K/2\}.$$

By definition of $\mathcal{D}_h$, we have $|A_q \setminus T_q| < t$. Our next goal is to find a prime $p' \leq K/q$ such that $\gcd(p', q) = 1$ and

$$|A_{p'q} \setminus T_q| \leq \frac{\log N}{1000 \log\log N} := B_0.$$

This follows by an averaging argument. Define

$$Y_q := \max\left(\sqrt{N/q},\ C_2\, t \frac{(\log\log N)^2}{\log N}\right), \qquad \mathcal{P}'_q := \left\{p' \text{ prime}:\ Y_q \leq p' \leq \min(2Y_q, K/q),\ (p', q) = 1\right\},$$

where $C_2 > 0$ is a large absolute constant to be fixed.

We first verify that $\mathcal{P}'_q \neq \emptyset$ and that $|\mathcal{P}'_q| \gg \frac{Y_q}{\log Y_q}$. It suffices to ensure that $2Y_q \leq K/q$. Equivalently, it is enough to check

$$2\sqrt{\frac{N}{q}} \leq \frac{K}{q} \qquad \text{and} \qquad 2C_2\, t\, \frac{(\log\log N)^2}{\log N} \leq \frac{K}{q}.$$

By the definition of S, for N sufficiently large we have

$$q \leq S \leq \frac{K^2}{4N},$$

and hence $2\sqrt{N/q} \leq K/q$ holds. Moreover,

$$2C_2\, t\, \frac{(\log\log N)^2}{\log N} \leq \frac{K}{q} \quad \Longleftrightarrow \quad q \leq \frac{K^3}{1800\, C_2\, N^2\, (\log\log N)^3},$$

which again follows from the definition of S (provided $C \geq 1800\, C_2$). Therefore $\mathcal{P}'_q \neq \emptyset$, and consequently $|\mathcal{P}'_q| \gg Y_q/\log Y_q$ as claimed.

Now observe that

$$\sum_{p'\in\mathcal{P}'_q} |A_{qp'} \setminus T_q| = \sum_{p'\in\mathcal{P}'_q} \big|\{n \in A_q \setminus T_q :\ p' \mid (n/q)\}\big|.$$

Since every $p' \in \mathcal{P}'_q$ satisfies $p' \geq Y_q \geq \sqrt{N/q}$ while $n/q \leq N/q$, each fixed $n \in A_q \setminus T_q$ can contribute to at most one prime $p' \in \mathcal{P}'_q$. Therefore

$$\sum_{p'\in\mathcal{P}'_q} |A_{qp'} \setminus T_q| \leq |A_q \setminus T_q| < t.$$

By the prime number theorem and the choice of Y_q (taking C_2 sufficiently large), we may ensure $|\mathcal{P}'_q| \geq t/B_0$, and hence by averaging there exists $p' \in \mathcal{P}'_q$ such that

$$p' \leq K/q \qquad \text{and} \qquad |A_{qp'} \setminus T_q| \leq B_0. \tag{3.3}$$

Note that $qp' \leq K$. Our next goal is to find r which is the product of at most two distinct primes at most S, with $qp'r \in [2K, 100K]$. If $S \geq 100K(qp')^{-1}$, then choose r to be a prime with $\gcd(r, qp') = 1$ in $[2K(qp')^{-1}, 100K(qp')^{-1}]$. Otherwise, let r_1 be a prime in $[S/2, S]$ with $\gcd(r_1, qp') = 1$, and if $qp'r_1 < 2K$ still, then let r_2 be a prime in $[2K(qp'r_1)^{-1}, 100K(qp'r_1)^{-1}]$ with $\gcd(qp'r_1, r_2) = 1$. Then let $r = r_1r_2$. Such r_1, r_2 exist because $S^2 \geq N^{1.98}$ which is much larger than K. Note that clearly,

$$|A_{qp'r} \setminus T_q| \leq |A_{qp'} \setminus T_q| \leq \frac{\log N}{1000 \log\log N}.$$

Let $\mathcal{P} = \{p \text{ prime} :\ 20\log N \leq p \leq 40 \log N\}$. Define

$$\mathcal{P}_q = \{p \in \mathcal{P} : \gcd(p, qp'r) = 1, A_{qp'rp} \subseteq T_q\}.$$

We claim that $|\mathcal{P}_q| \geq 0.99|\mathcal{P}|$ for N large. Indeed, if $p \notin \mathcal{P}_q$, then either $p \mid qp'r$ (excluding at most 4 primes), or else there exists $n \in A_{qp'rp} \setminus T_q$. In the latter case, $n \in A_{qp'r} \setminus T_q$ and $p \mid (n/qp'r)$. Since $qp'r \geq 2K$ and $n \leq N$, we have $n/qp'r \leq N/(2K) = 5 \cdot 10^6 \log N$. For N large, $5 \cdot 10^6 \log N < (20 \log N)^2$, so such an integer cannot have two distinct prime factors in $[20 \log N, 40 \log N]$. Hence each fixed $n \in A_{qp'r} \setminus T_q$ rules out at most one prime $p \in \mathcal{P}$. Therefore,

$$|\mathcal{P} \setminus \mathcal{P}_q| \leq 4 + |A_{qp'r} \setminus T_q| \ll 1 + B_0 \leq 0.01|\mathcal{P}|$$

for N large, since $|\mathcal{P}| \geq \log N/\log\log N$ while $B_0 = \log N/(1000 \log\log N)$.

Note that $qp'r \geq K$, and thus there is a unique multiple x_q of $qp'r$ in I_h, if it exists. Let $p \in \mathcal{P}_q$. Because $K \leq qp'rp \leq 4000K \log N \leq N/2000$, there is a multiple of n of $qp'rp$ that is S-smooth in the range $[N/2, N]$. We also require the multiple to have sufficiently few prime divisors; here we use that $\widetilde{\Omega}(n) \leq 5 \log\log N$ implies that $\Omega(qp'rp) \leq 5 \log\log N + 4$. Because $n \in A_{qp'rp} \subseteq T_q$, we know that $|h_n| < K/2$, there is $x \in I_h$ with $n|x$. Because $qp'r|x$, we know that $x = x_q$, and thus $p|x$ too.

Now, for any $q_1, q_2 \in \mathcal{D}_h$, we know that $|\mathcal{P}_{q_1} \cap \mathcal{P}_{q_2}| \geq 0.8|\mathcal{P}|$. By the reasoning in the above paragraph, we deduce that $\prod_{p\in\mathcal{P}_{q_1}\cap\mathcal{P}_{q_2}} p|x_{q_1}, x_{q_2}$, and thus $x_{q_1} = x_{q_2}$ because

$$\prod_{p\in\mathcal{P}_{q_1}\cap\mathcal{P}_{q_2}} p > (20 \log N/(\log\log N))^{5\log N/(\log\log N)} > N.$$

Thus, all x_q are equal for $q \in \mathcal{D}_h$, and thus $[\mathcal{D}_h]$ divides the common x_q.

Minor arcs, finishing the proof: We are now in position to complete the proof. Consider all $h \in (-Q/2, Q/2]$ such that $\mathcal{D}_h = \mathcal{D}$. As $[\mathcal{D}_h] | x$ for an element $x \in I_h$, we have that the number of such h is bounded by

$$(K+1) \cdot \frac{[\mathcal{Q}_A]}{[\mathcal{D}]} \le N \cdot \prod_{q \in \mathcal{Q}_A \setminus \mathcal{D}} q \le N^{|\mathcal{Q}_A \setminus \mathcal{D}|+1}. \tag{3.4}$$

As we have restricted attention to $|h| > M/2 \ge K/2$, we have that $\mathcal{D}_h \neq \mathcal{Q}_A$. Therefore the total contribution over minor arcs, using (3.2) and (3.4) and is bounded by

$$\frac{1}{Q} \sum_{\mathcal{D} \subsetneq \mathcal{Q}_A} N^{|\mathcal{Q}_A \setminus \mathcal{D}|+1} \cdot N^{-10|\mathcal{Q}_A \setminus \mathcal{D}|} \le \frac{1}{Q} \cdot \sum_{s \ge 1} N^{s+1} \cdot N^s \cdot N^{-10s} \le 2/(QN).$$

This completes the proof. □

Now we present the strengthened version of [3, Lemma 4.1].

Lemma 3.3. *There exist absolute constants $c_1 > 0$ and N_1 such that the following holds for all $N \ge N_1$. Let*

$$S := c_1 \frac{N}{(\log N)^3 (\log\log N)^3}.$$

Assume that t_0 is S-smooth. If $t/3 \le s \le t$, then there exists a finite set $A \subseteq [N/16, N] \cap \mathbb{N}$ such that $\sum_{n \in A} \frac{1}{n} = s/t$.

Proof. Apply Proposition 3.1 with the following choices of parameters:

$$M = N/16, K = 10^{-7} N (\log N)^{-1}, \quad \text{and} \quad S = c_1 N (\log N)^{-3} (\log\log N)^{-3},$$

for sufficiently small c_1. These satisfy the hypotheses of Proposition 3.1. By [3, Lemma 3.3] we know that $|A| \ge 0.89N$, so $R(A) \ge \sum_{n=0.11N}^{N} \frac{1}{n} \ge 2$. Also, $R(A) \le 3$. Let $p_n = \frac{s/t}{R(A)}$ for all $n = N/16, \ldots, N$ so that $\mathbb{E}[R(B)] = s/t$. Notice that we clearly have $1/9 \le p_n \le 1/2$. Finally, it suffices to check that there is an integer x with $x/Q = s/t$, i.e., $t | Q$. This holds exactly because t is S-smooth. □

4. Main Upper Bound

Theorem 4.1. *There exist absolute constants $c_0 > 0$ and $N_0 \in \mathbb{N}$ such that for all $N \ge N_0$,*

$$\delta(N) \le \exp\left(-c_0 \frac{N}{(\log N)^3 (\log\log N)^3} \right).$$

Proof. Let $N \ge N_0$ be sufficiently large, where N_0 will be chosen at the end. Define

$$S := c_1 \frac{N}{(\log N)^3 (\log\log N)^3}, \qquad S_0 := \lfloor S \rfloor, \qquad t := \mathrm{lcm}(1, 2, \ldots, S_0), \qquad s := t - 1.$$

Since every prime power divisor of t is at most $S_0 \le S$, the integer t is S–smooth. Also, once $S_0 \ge 2$ we have $t \ge 2$, hence $t/3 \le t - 1 = s \le t$.

Applying Lemma 3.3 to these parameters (s, t), we obtain a set

$$A \subset [N/16, N] \cap \mathbb{N} \subseteq \{1, 2, \ldots, N\}$$

such that

$$\sum_{n \in A} \frac{1}{n} = \frac{s}{t} = 1 - \frac{1}{t}.$$

In particular $\sum_{n\in A}\frac{1}{n}<1$, so A is admissible for the definition of $\delta(N)$. Therefore

$$\delta(N) \;\le\; \left|1-\sum_{n\in A}\frac{1}{n}\right| = \frac{1}{t}. \tag{4.1}$$

It remains to lower bound t. By (2.1) and (2.2), provided $S_0 \ge x_0$ we have

$$\log t = \log \operatorname{lcm}(1,2,\dots,S_0) = \psi(S_0) \ge c_\psi S_0.$$

If additionally $S_0 \ge 2$, then $S_0 \ge S/2$, hence

$$\log t \ge c_\psi S_0 \ge \frac{c_\psi}{2} S.$$

Consequently,

$$\frac{1}{t} \le \exp\Big(-\frac{c_\psi}{2}S\Big) = \exp\bigg(-\frac{c_\psi c_1}{2}\,\frac{N}{(\log N)^3(\log\log N)^3}\bigg).$$

Combining this with (4.1) gives the claimed bound with $c_0 := \frac{c_\psi c_1}{2}$.

Finally, choose N_0 large enough so that simultaneously: (i) $N_0 \ge N_1$ (to apply Lemma 3.3), (ii) $S_0 = \lfloor S\rfloor \ge x_0$ (to apply (2.2)), and (iii) $S_0 \ge 2$ (so that $t \ge 2$ and $S_0 \ge S/2$). This completes the proof. □

References

[1] T. F. Bloom, Erdős Problem #311, https://www.erdosproblems.com/311, accessed 2026-01-12.

[2] P. Erdős, R. Graham, Old and new problems and results in combinatorial number theory. Monographies de L'Enseignement Mathematique (1980).

[3] Y. P. Liu, M. Sawhney, On further questions regarding unit fractions. arXiv preprint arXiv:2404.07113v1. https://arxiv.org/abs/2404.07113v1

School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, P. R. China
Email address: tang_quanyu@163.com