Generalized harmonic sums have arbitrarily large prime factors

Wouter van Doorn

The purpose of this note is three-fold. First of all, it contains a proof of Theorem 13 in [1]. Second, it serves as the solution of a generalization of the first half of Erdős problem #291 [2]. In an equivalent form, this problem states that the denominator of the n-th partial harmonic sum is infinitely often not equal to the least common multiple of the first n integers. And thirdly, it provides a document that was given to the automated theorem proving program Aristotle from Harmonic, in order to get the results formalized into Lean. The formalization we obtained this way can be found at the author's GitHub page [3]. We note that this formalization uses two prime number theorem type results (found below as Lemma 2 and Lemma 3) as axioms. These results have, by the time you are reading this, most likely already been formalized by the PNT+ Project [4]. Okay, on to the mathematics, which are often taken close to verbatim from Section 2.4 and Section 2.5 of [1].

Let $r_1, r_2, \dots$ be a bounded sequence of non-zero integers, let L_n denote the least common multiple of $1, 2, \dots, n$ and define $X_n = L_n \sum_{i=1}^{n} \frac{r_i}{i}$. Let $m \geq 4$ be any integer larger than $\max_i |r_i|$, and let $2 = p_1 < p_2 < \dots < p_z < m$ be the sequence of primes smaller than m. The goal is to prove that there exists an n such that X_n has a prime divisor larger than or equal to m. From this result we will then deduce that, if the sequence of r_i is periodic, then there are infinitely many n for which $\gcd(X_n, L_n) > 1$. However, before we can find large prime divisors of X_n, we first have to find an interval where $|X_n|$ itself is large.

Let $\tilde{m}$ be any integer larger than $20m^{2z}$ such that $\tilde{m}$ has a prime divisor q_0 larger than m^{2z-1}. Furthermore define the interval

$$I := [\tilde{m} - m^{2z-1}, \tilde{m} + m^{2z-1}),$$

and divide it into the sub-intervals

$$J_1 := [\tilde{m} - m^{2z-1}, \tilde{m})$$

and

$$J_2 := [\tilde{m}, \tilde{m} + m^{2z-1}).$$

Lemma 1. *Either $|X_n| > n^z$ for all $n \in J_1$ or $|X_n| > n^z$ for all $n \in J_2$.*

Proof. Without loss of generality we may assume that there exists an integer $w \in J_1$ with $|X_w| \leq w^z$. Let $w + k$ be an integer in J_2 and note

$$k < (\tilde{m} + m^{2z-1}) - (\tilde{m} - m^{2z-1}) = 2m^{2z-1}.$$

Our goal is to prove $|X_{w+k}| > (w+k)^z$, but we first need a few technical lemmas.

Lemma 2. *For all $m \in \mathbb{N}$ we have $m^{2z} < e^{2.52m}$.*

Lemma 3. *For all $k \in \mathbb{N}$ we have the lower bound $L_{w+k} > 2^{w+k}$.*

Lemma 4. *For all $k \in \mathbb{N}$ with $w + k \in J_2$ we have the lower bound*

$$\left| \sum_{i=w+1}^{w+k} \frac{r_i}{i} \right| \geq \frac{1}{(w+k)^k}.$$

Lemma 5. *For all $k \in \mathbb{N}$ with $w + k \in J_2$ we have the inequality:*

$$\frac{2^{w+k}}{(w+k)^k} - (w+k)^k w^z > (w+k)^z.$$

Lemma 2 and Lemma 3 follow from [5, Corollary 1] and [5, Theorem 10] respectively. As mentioned before, our Lean proof uses these two lemmas as axioms.

Proof of Lemma 4. The sum $\displaystyle\sum_{i=w+1}^{w+k} \frac{r_i}{i}$ can be written as a fraction with denominator equal to $L := \operatorname{lcm}(w+1, \ldots, w+k)$, which is trivially upper bounded by $(w+k)^k$. So to prove that the estimate we want to show holds, it suffices to show that the left-hand side is non-zero. Note that

$$\tilde{m} \leq w + k < \tilde{m} + m^{2z-1} < \tilde{m} + q_0.$$

So in the sum $L \displaystyle\sum_{i=w+1}^{w+k} \frac{r_i}{i}$, every term is divisible by q_0, except for the term corresponding to $i = \tilde{m}$. The term corresponding to $i = \tilde{m}$ is not divisible by q_0 as $0 < |r_{\tilde{m}}| < q_0$. Since the sum is then not divisible by q_0, it is certainly non-zero, which means $\displaystyle\sum_{i=w+1}^{w+k} \frac{r_i}{i}$ is non-zero as well. $\square$

Proof of Lemma 5. We calculate, using the fact that $\frac{x}{\log(x)}$ is an increasing function of x for $x \geq 3$, applying the inequalities $w + k \geq \tilde{m} \geq 20m^{2z}$ and

$m^{2z} < e^{2.52m}$, and making use of the bounds $m \geq \max(4, z)$ and $4m^{2z-1} > 2k$;

$$\begin{aligned}
\frac{w+k}{\log(w+k)} &\geq \frac{20m^{2z}}{\log(20m^{2z})} \\
&> \frac{20m^{2z}}{\log(20e^{2.52m})} \\
&= \frac{20m^{2z}}{\log(20) + 2.52m} \\
&> \frac{20m^{2z}}{3.3m} \\
&> 6m^{2z-1} \\
&> 2m + 5.8m^{2z-1} \\
&> 2z + \frac{4m^{2z-1}}{\log(2)} \\
&> \frac{1}{\log(w+k)} + \frac{z}{\log(2)} + \frac{2k}{\log(2)}.
\end{aligned}$$

When we multiply by $\log(w+k)$ and take 2 to the power of both sides, we obtain

$$\begin{aligned}
2^{w+k} &> 2(w+k)^{2k+z} \\
&> (w+k)^{2k+z} + (w+k)^{k+z} \\
&> (w+k)^{2k} w^z + (w+k)^{k+z}.
\end{aligned}$$

Dividing by $(w+k)^k$ and rearranging gives the desired inequality. $\square$

Combining all these lemmas lets us finish the proof that $|X_{w+k}| > (w+k)^z$ for an arbitrary integer $w + k \in J_2$. Indeed,

$$\begin{aligned}
|X_{w+k}| &= \left| L_{w+k} \sum_{i=1}^{w+k} \frac{r_i}{i} \right| \\
&= \left| \frac{L_{w+k}}{L_w} X_w + L_{w+k} \sum_{i=w+1}^{w+k} \frac{r_i}{i} \right| \\
&\geq L_{w+k} \left| \sum_{i=w+1}^{w+k} \frac{r_i}{i} \right| - \frac{L_{w+k}}{L_w} |X_w| \\
&\geq \frac{2^{w+k}}{(w+k)^k} - (w+k)^k w^z \\
&> (w+k)^z.
\end{aligned}$$

$\square$

Now we set set $I_0 = J_1$ if $|X_n| > n^z$ holds true for all $n \in J_1$, or else set $I_0 = J_2$. We will then prove the following theorem.

Theorem 1. *There exists an integer $n \in I_0$ for which X_n is divisible by a prime larger than or equal to m.*

Proof. Define $e_i(x)$ to be the largest power of p_i that divides x. With this notation, $p_1^{e_1(X_n)} \cdots p_z^{e_z(X_n)}$ is the prime decomposition of the largest divisor $d(n)$ of X_n which consists only of primes smaller than m. The goal is to find an $n \in I_0$ with $d(n) \leq n^z$, which would imply that X_n has a prime divisor larger than or equal to m.

We shall construct a sequence $n_1 < n_2 < \cdots < n_{z+1}$ of integers contained in I_0, such that either $d(n_j) \leq n_j^z$ for some j with $1 \leq j \leq z$, or for n_{z+1} we have that $p_i^{e_i(X_{n_{z+1}})} \leq n_{z+1}$ holds for all i with $1 \leq i \leq z$, implying $d(n_{z+1}) \leq n_{z+1}^z$.

To start off, define n_1 to be the smallest integer in I_0. Now, once we have defined n_j for some j with $1 \leq j \leq z$, if $d(n_j) \leq n_j^z$, we are done. So for the rest of this proof we are free to assume that, after we have defined n_j, the inequality $d(n_j) > n_j^z$ holds. This implies in particular that there exists a $\sigma(j) \in \{1, 2, \ldots, z\}$ with $p_{\sigma(j)}^{e_{\sigma(j)}(X_{n_j})} > n_j$. Of course, there can be more than one such prime. Just pick, say, the smallest.

Then let $p_{\sigma(j)}^{k_j}$ be the largest power of $p_{\sigma(j)}$ smaller than or equal to m^{2z-2j} and set n_{j+1} equal to the smallest integer larger than n_j such that

$$e_{\sigma(j)}(n_{j+1}) - e_{\sigma(j)}(r_{n_{j+1}}) \geq k_j.$$

Further, define the half-open interval

$$I_j := [n_{j+1}, n_{j+1} + p_{\sigma(j)}^{k_j}).$$

Then we claim that the intervals I_j form a decreasing sequence.

Lemma 6. *We have $I_0 \supset I_1 \supset I_2 \supset \cdots \supset I_z$.*

Proof. Since $I_j = [n_{j+1}, n_{j+1} + p_{\sigma(j)}^{k_j})$ for $j \geq 1$, we note that the statement $I_{j-1} \supset I_j$ for $j \geq 2$ is equivalent to the following two inequalities:

$$\begin{gathered} n_j \leq n_{j+1} \\ n_{j+1} + p_{\sigma(j)}^{k_j} \leq n_j + p_{\sigma(j-1)}^{k_{j-1}} \end{gathered}$$

While for $I_0 \supset I_1$ the second inequality gets replaced by

$$n_2 + p_{\sigma(1)}^{k_1} \leq n_1 + m^{2z-1}.$$

First of all, as n_{j+1} is defined as the smallest integer larger than n_j with some property, the inequality $n_j \leq n_{j+1}$ is trivial. Secondly, recall that n_{j+1} is defined as the smallest integer larger than n_j with

$$e_{\sigma(j)}(n_{j+1}) - e_{\sigma(j)}(r_{n_{j+1}}) \geq k_j.$$

And as

$$e_{\sigma(j)}(r_{n_{j+1}}) \leq \left\lfloor \frac{\log(m-1)}{\log(p_{\sigma(j)})} \right\rfloor,$$

we deduce

$$n_{j+1} \leq n_j + (m-1)p_{\sigma(j)}^{k_j}.$$

Lastly, we look for bounds on $p_{\sigma(j)}^{k_j}$. Again we have a trivial bound

$$p_{\sigma(j)}^{k_j} \leq m^{2z-2j},$$

because $p_{\sigma(j)}^{k_j}$ is defined as the largest power of $p_{\sigma(j)}$ smaller than or equal to m^{2z-2j}. On the other hand, there is always a power of $p_{\sigma(j)}$ between two consecutive powers of m since $p_{\sigma(j)} < m$. So $p_{\sigma(j)}^{k_j}$ must be larger than $m^{2z-2j-1}$. By putting all these inequalities together we can prove $I_{j-1} \supset I_j$, for all $j \in \{2, \ldots, z\}$:

$$\begin{aligned} n_{j+1} + p_{\sigma(j)}^{k_j} &\leq n_j + (m-1)p_{\sigma(j)}^{k_j} + p_{\sigma(j)}^{k_j} \\ &= n_j + mp_{\sigma(j)}^{k_j} \\ &\leq n_j + m^{2z+1-2j} \qquad (1) \\ &= n_j + m^{2z-1-2(j-1)} \\ &< n_j + p_{\sigma(j-1)}^{k_{j-1}}. \end{aligned}$$

To prove $I_0 \supset I_1$, use the above reasoning up to and including equation (1) with $j = 1$. ∎

Lemma 7. *For all $n \in I_j$ we have $p_{\sigma(j)}^{e_{\sigma(j)}(X_n)} \leq n$.*

Proof. Let n be any integer in $I_j \subset I_0$. We will then write X_n as a sum of four distinct terms.

$$\begin{aligned} X_n &= L_n \sum_{i=1}^{n} \frac{r_i}{i} \\ &= \sum_{i=1}^{n_j} \frac{L_n r_i}{i} + \sum_{i=n_j+1}^{n_{j+1}-1} \frac{L_n r_i}{i} + \frac{L_n r_{n_{j+1}}}{n_{j+1}} + \sum_{i=n_{j+1}+1}^{n} \frac{L_n r_i}{i} \end{aligned}$$

By assumption, X_{n_j} is divisible by a power of $p_{\sigma(j)}$ larger than n_j, hence we obtain

$$e_{\sigma(j)}\left(\frac{L_n X_{n_j}}{L_{n_j}}\right) \geq e_{\sigma(j)}(L_n) + 1 \geq e_{\sigma(j)}(L_n) - k_j + 1$$

for the first term.

As for the second and third terms, by the definition of n_{j+1} we know that for every $i \in [n_j + 1, n_{j+1} - 1]$ we have

$$e_{\sigma(j)}\left(\frac{L_n r_i}{i}\right) \geq e_{\sigma(j)}(L_n) - k_j + 1,$$

while

$$e_{\sigma(j)}\left(\frac{L_n r_{n_{j+1}}}{n_{j+1}}\right) \leq e_{\sigma(j)}(L_n) - k_j.$$

Finally, since $e_{\sigma(j)}(n_{j+1}) \geq k_j$ and $n < n_{j+1} + p_{\sigma(j)}^{k_j}$, we have $e_{\sigma(j)}(i) < k_j$ for all $i \in [n_{j+1} + 1, n]$, hence

$$e_{\sigma(j)}\left(\frac{L_n r_i}{i}\right) \geq e_{\sigma(j)}(L_n) - k_j + 1.$$

Combining the above estimates we see that there is exactly one term in the sum for X_n that is not divisible by $p_{\sigma(j)}^{e_{\sigma(j)}(L_n)-k_j+1}$, and we conclude that the largest power of $p_{\sigma(j)}$ that divides X_n is at most $p_{\sigma(j)}^{e_{\sigma(j)}(L_n)-k_j} \leq n$. $\square$

Now we may finish the proof. First off, all the $p_{\sigma(j)}$ have to be distinct, since $p_{\sigma(i)}^{e_{\sigma(i)}(X_{n_i})} > n_i$, while Lemma 7 shows that if $i > j$, then for all $n \in I_{i-1} \subset I_j$ it holds true that $p_{\sigma(j)}^{e_{\sigma(j)}(X_n)} \leq n$. In other words, $(\sigma(1), \sigma(2), \ldots, \sigma(z))$ is a permutation of $(1, 2, \ldots, z)$. Secondly, since our intervals form a nesting sequence, for $n_{z+1} \in I_z \subset I_j$ we have $p_{\sigma(j)}^{e_{\sigma(j)}(X_{n_{z+1}})} \leq n_{z+1}$ for all j with $1 \leq j \leq z$. We conclude that

$$d(n_{z+1}) = \prod_{j=1}^{z} p_j^{e_j(X_{n_{z+1}})} = \prod_{j=1}^{z} p_{\sigma(j)}^{e_{\sigma(j)}(X_{n_{z+1}})} \leq \prod_{j=1}^{z} n_{z+1} = n_{z+1}^z,$$

and the theorem is proved. $\square$

Corollary 1. *If the sequence of numerators $r_1, r_2, \ldots$ is periodic, then we have $\limsup \gcd(X_b, L_b) = \infty$. In particular, there are in that case infinitely many b for which $\gcd(X_b, L_b) > 1$.*

Proof. Assume that the period of the sequence of r_i is equal to t. That is, we have $r_i = r_{i+t}$ for all $i \in \mathbb{N}$. Let $N \in \mathbb{R}$ be arbitrary, and let $m \geq 4$ be any integer larger than $\max(|r_1|, |r_2|, \ldots, |r_t|, t, N)$. We will then show that there exists a b such that $\gcd(X_b, L_b) \geq m > N$. Since N is arbitrary, this shows $\limsup \gcd(X_b, L_b) = \infty$.

By Theorem 1 there exists an n such that X_n is divisible by a prime p with $p \geq m$. If $n \geq p$, then p also divides L_n, so that with $b = n$ we have $\gcd(X_b, L_b) \geq p \geq m$ and we are done. We may therefore assume $1 \leq n < p$, in which case p does not divide L_n. We then claim that $b = np^{\varphi(t)}$ works, where φ represents Euler's totient function.

Note that with this value of b we have that $p^{\varphi(t)}$ divides L_b. And hence, in order to calculate $X_b \pmod p$, we have $\frac{L_b r_i}{i} \equiv 0 \pmod p$ unless $p^{\varphi(t)}$ divides i. That is, if $\frac{L_b r_i}{i} \not\equiv 0 \pmod p$, then $i = jp^{\varphi(t)}$ for some j with $1 \leq j \leq n$.

As $p \geq m > t$ we have $\gcd(p, t) = 1$, so that $jp^{\varphi(t)} \equiv j \pmod t$ by Euler's Theorem. In particular, we have $r_{jp^{\varphi(t)}} = r_j$ since the sequence of $r_1, r_2, \ldots$ is by assumption periodic with period t. We therefore conclude

$$\begin{aligned}
X_b &= L_b \sum_{i=1}^{b} \frac{r_i}{i} \\
&= \sum_{i=1}^{b} \frac{L_b r_i}{i} \\
&\equiv \sum_{j=1}^{n} \frac{L_b r_{jp^{\varphi(t)}}}{jp^{\varphi(t)}} && \pmod p \\
&\equiv \frac{L_b}{p^{\varphi(t)}} \sum_{j=1}^{n} \frac{r_j}{j} && \pmod p \\
&\equiv \frac{L_b}{L_n p^{\varphi(t)}} L_n \sum_{j=1}^{n} \frac{r_j}{j} && \pmod p \\
&\equiv \frac{L_b}{L_n p^{\varphi(t)}} X_n && \pmod p \\
&\equiv 0 && \pmod p,
\end{aligned}$$

so that both X_b and L_b are divisible by p. We deduce $\gcd(X_b, L_b) \geq p \geq m > N$, as desired. □

References

[1] W. van Doorn, *On the non-monotonicity of the denominator of generalized harmonic sums.* arXiv:2411.03073, 2024.

[2] T. F. Bloom, Erdős Problem #291, https://www.erdosproblems.com, accessed 2026-02-05.

[3] ErdosProblem#291. `https://github.com/Woett/Lean-files/blob/main/ErdosProblem%23291.lean`.

[4] PrimeNumberTheoremAnd. `https://github.com/AlexKontorovich/PrimeNumberTheoremAnd`.

[5] J. B. Rosser, L. Schoenfeld, *Approximate formulas for some functions of prime numbers*. Illinois J. Math. 6, no. 1, 64–94, 1962.