

## On unit fractions with denominators in short intervals

by

ERNEST S. CROOT III (Berkeley, CA)

*Dedicated to the memory of Paul Erdős*

**I. Introduction.** Let  $X_k$  denote the set

$$\left\{ \{x_1, \dots, x_k\} : \sum_{j=1}^k \frac{1}{x_j} = 1, 0 < x_1 < \dots < x_k \right\}.$$

Erdős and Graham (see [3] and [4]) asked the following questions:

1. Is it true that

$$\max\{x_1 : \{x_1, \dots, x_k\} \in X_k\} \sim \frac{k}{e-1}?$$

Trivially, it is less than or equal to  $(1 + o(1))k/(e-1)$ , so all one needs to show is a lower bound of size  $(1 + o(1))k/(e-1)$ .

2. Is it true that

$$\min\{x_k - x_1 : \{x_1, \dots, x_k\} \in X_k\} \sim k?$$

(Note: These two questions were misstated in [3].)

In [6] Greg Martin proves that there exist  $\sim (e-1)x/e$  integers  $\leq x$  whose sum of reciprocals equals 1 (actually he proves a more general result, which applies to expansions of positive rationals).

This result implies that

$$\min\{x_k : \{x_1, \dots, x_k\} \in X_k\} \sim \frac{ek}{e-1},$$

which improves upon his earlier work in [5]. This result would also follow from the affirmative answer to questions 1 and 2 above; however, Martin's result cannot be applied to solve these questions, since his result gives no information about  $x_1$  (the smallest denominator in such representation).

---

2000 *Mathematics Subject Classification*: Primary 11P99.

In this paper we will prove the following theorem, which solves these questions of Erdős and Graham for infinitely many  $k$ .

**MAIN THEOREM.** *Suppose that  $r > 0$  is any given rational number. Then, for all  $N > 1$ , there exist integers  $x_1, \dots, x_k$ , with*

$$N < x_1 < \dots < x_k \leq \left( e^r + O_r \left( \frac{\log \log N}{\log N} \right) \right) N$$

such that

$$r = \frac{1}{x_1} + \dots + \frac{1}{x_k}.$$

Moreover, the error term  $O_r(\log \log N / \log N)$  is best possible.

We will now discuss the idea of the proof of the Main Theorem. Let  $c > 1$  be the smallest real number such that

$$r \leq \sum_{N < n < cN} \frac{1}{n} \leq r + \frac{1}{cN}.$$

Using the fact that  $\sum_{1 \leq n \leq t} 1/n = \log t + \gamma + O(1/t)$  one can show that  $c = e^r + O_r(1/N)$ . Now suppose

$$(1.1) \quad \frac{u}{v} = \sum_{N < n < cN} \frac{1}{n}, \quad \text{where } \gcd(u, v) = 1.$$

If we had  $u/v = r$ , then we would have proved our theorem for this instance of  $r$  and  $N$ , because  $c = e^r + O_r(1/N)$  is well within the error of  $O_r(\log \log N / \log N)$  claimed by our theorem. Unfortunately, for large  $N$  it will not be the case that  $u/v = r$ .

To prove the theorem, we first will use a proposition which says that we can remove terms from the sum in (1.1), call them  $1/d_1, \dots, 1/d_l$ , so that if

$$\frac{u'}{v'} = \frac{u}{v} - \left\{ \frac{1}{d_1} + \dots + \frac{1}{d_l} \right\} = \sum_{\substack{N < n < cN \\ n \neq d_1, \dots, d_l}} \frac{1}{n}, \quad \text{where } \gcd(u', v') = 1,$$

then all the prime power factors of  $v'$  are  $\leq N^{1/5}$ ; moreover, we will have

$$\frac{\log \log N}{\log N} \ll_r \frac{1}{d_1} + \dots + \frac{1}{d_l} \ll_r \frac{\log \log N}{\log N}.$$

We will then couple this with another proposition which says that if  $s$  is some rational number whose denominator has all its prime power factors  $\leq M^{1/4-\varepsilon}$  where  $0 < \varepsilon < 1/8$ , and if  $s \gg \log \log \log M / \log M$ , then there are integers  $M < n_1 < \dots < n_k < e^{(v(\varepsilon)+o(1))s} M$ , where  $v(\varepsilon)$  is some constant depending on  $\varepsilon$ , such that

$$s = \frac{1}{n_1} + \dots + \frac{1}{n_k}.$$

The way we use this second proposition is by letting  $M = cN$  and setting  $s = r - u'/v'$ , so that

$$\frac{\log \log M}{\log M} \ll_r s \ll_r \frac{\log \log M}{\log M}.$$

Now, all the prime power factors of the denominator of  $s$  will be  $\leq N^{1/5}$  (when  $N$  is sufficiently large). Thus, the hypotheses of this second proposition are met with  $\varepsilon = 1/20$ , and so there exist  $n_1, \dots, n_k$  such that

$$r = s + \frac{u'}{v'} = \sum_{\substack{N < n < cN \\ n \neq d_1, \dots, d_l}} \frac{1}{n} + \sum_{i=1}^k \frac{1}{n_i}.$$

All the denominators of these unit fractions will be no larger than

$$e^{(v(20)+o(1))s} M = e^{(v(20)+o(1))s+r} N = \left( e^r + O_r \left( \frac{\log \log N}{\log N} \right) \right) N,$$

and will all be greater than  $N$ .

The way we will prove that the error term  $O_r(\log \log N/\log N)$  is best possible is by showing that if

$$r = \frac{1}{x_1} + \dots + \frac{1}{x_k}, \quad 2 \leq x_1 < \dots < x_k \text{ are integers,}$$

then none of the  $x_i$ 's can be divisible by a prime  $p > x_k/\log x_k$  (this idea appears in [2], [3], and [6]). It will turn out that this forces

$$\frac{x_k}{x_1} > e^r \left( 1 + \frac{(r + o(1)) \log \log x_k}{\log x_k} \right),$$

thus finishing the proof of the Main Theorem.

**II. Smooth numbers.** In order to even state, let alone prove, the propositions and lemmas needed to prove the Main Theorem, we will need to introduce some notation and definitions concerning smooth numbers. We say that a number  $n$  is *y-smooth* if all of its prime factors are less than or equal to  $y$ , and we define the usual smooth number counting function as follows:

$$\begin{aligned} \psi(N, y) &:= \#\{n \leq N : n \text{ is } y\text{-smooth}\} \\ &= \#\{n \leq N : p \mid n, p \text{ prime} \Rightarrow p \leq y\}. \end{aligned}$$

Define

$$S(N, y) := \{n \leq N : p^a \mid n, p \text{ prime} \Rightarrow p^a \leq y\},$$

and let

$$\psi'(N, y) = |S(N, y)|,$$

the number of elements in  $S(N, y)$ .

In later sections we will need various estimates concerning the  $\psi'(N, y)$  and  $\psi(N, y)$  functions, and we will use the following lemma to obtain them.

LEMMA 1 (N. G. de Bruijn). *For any fixed  $\varepsilon < 3/5$ , uniformly in the range*

$$y \geq 2, \quad 1 \leq u \leq \exp\{(\log y)^{3/5-\varepsilon}\},$$

*we have*

$$\psi(N, y) = N\rho(u) \left\{ 1 + O\left(\frac{\log(u+1)}{\log y}\right) \right\},$$

*where  $u = \log N/\log y$  and  $\rho(u)$  is the unique continuous solution to the differential-difference equation*

$$\begin{cases} \rho(u) = 1 & \text{if } 0 \leq u \leq 1, \\ u\rho'(u) = -\rho(u-1) & \text{if } u > 1. \end{cases}$$

For a proof of this lemma, see [1]. We can deduce the same estimate for the function  $\psi'(N, y)$  by using the following lemma.

LEMMA 2.

$$\sum_{\substack{mp^a \leq L \\ p^a \geq y, a \geq 2 \\ p \text{ prime}}} \frac{1}{mp^a} = O\left(\frac{\log L}{\sqrt{y}}\right).$$

*Proof.*

$$\sum_{\substack{mp^a \leq L \\ p^a \geq y, a \geq 2 \\ p \text{ prime}}} \frac{1}{mp^a} < \sum_{n \geq \sqrt{y}} \sum_{j=2}^{\infty} \sum_{m \leq L} \frac{1}{n^j m} \ll \log L \sum_{n \geq \sqrt{y}} \frac{1}{n^2} \ll \frac{\log L}{\sqrt{y}}.$$

From these last two lemmas we deduce that

$$\begin{aligned} (2.1) \quad \psi'(N, y) &= \psi(N, y) - O\left(N \sum_{\substack{mp^a \leq N \\ p^a \geq y, a \geq 2 \\ p \text{ prime}}} \frac{1}{mp^a}\right) \\ &= \psi(N, y) - O\left(\frac{N \log N}{\sqrt{y}}\right). \end{aligned}$$

Combining this with the previous two lemmas, we have the following final result of this section.

LEMMA 3. *If  $c, u \ll 1$  and  $N \gg_{c,u} 1$ , then*

$$\sum_{\substack{N < n < cN \\ n \in S(N, N^{1/u})}} \frac{1}{n} = \rho(u) \log c + O_u\left(\frac{1}{\log N}\right).$$

*Proof.* From Lemma 1, Lemma 2, and (2.1) we have the following chain of equalities:

$$\begin{aligned} \sum_{\substack{N < n < cN \\ n \in S(N, N^{1/u})}} \frac{1}{n} &= \left\{ \sum_{\substack{N < n < cN \\ n \text{ is } N^{1/u}\text{-smooth}}} \frac{1}{n} \right\} - O\left(\frac{\psi(N, N^{1/u}) - \psi'(N, N^{1/u})}{N}\right) \\ &= \sum_{\substack{N < n < cN \\ n \text{ is } n^{1/u}\text{-smooth}}} \frac{1}{n} - O\left(\sum_{\substack{N^{1/u} < p < (cN)^{1/u} \\ p \text{ prime}}} \sum_{N/p < m < cN/p} \frac{1}{mp}\right) \\ &\quad - O\left(\frac{\log(cN)}{N^{1/(2u)}}\right) \\ &= \varrho(u) \log c + O\left(\frac{1}{\log N} + \frac{\pi((cN)^{1/u})}{N^{1/u}} + \frac{\log(cN)}{N^{1/(2u)}}\right) \\ &= \varrho(u) \log c + O_u\left(\frac{1}{\log N}\right). \end{aligned}$$

**III. Proof of the Main Theorem.** To prove the Main Theorem we will require the following two propositions, which are the same as those mentioned in the introduction.

PROPOSITION 1. *Let  $c > 1$ . Then, for all  $N$  sufficiently large, there exist integers  $d_1, \dots, d_l$  with  $N < d_1 < \dots < d_l < cN$  such that if*

$$(3.1) \quad \frac{f}{g} = \sum_{\substack{N < n < cN \\ n \neq d_1, \dots, d_l}} \frac{1}{n},$$

*then all the prime power factors of  $g$  are  $\leq N^{1/5}$ , and*

$$(3.2) \quad \frac{\log \log N}{\log N} \ll_c \frac{1}{d_1} + \dots + \frac{1}{d_l} \ll_c \frac{\log \log N}{\log N}.$$

PROPOSITION 2. *Suppose  $0 < \varepsilon < 1/8$  and  $A$  and  $B$  are positive integers, where  $\gcd(A, B) = 1$ , all the prime power divisors of  $B$  are  $\leq M^{1/4-\varepsilon}$ , and  $\log \log \log M / \log M \ll A/B \leq 1$ . Select  $c(M) > 0$  so that*

$$2\frac{A}{B} \leq \sum_{\substack{M \leq n \leq c(M)M \\ n \in S(c(M)M, M^{1/4-\varepsilon})}} \frac{1}{n} < 2\frac{A}{B} + \frac{1}{c(M)M}.$$

*Then, for all  $M$  sufficiently large, there exist integers  $n_1, \dots, n_k$  with  $M \leq n_1 < \dots < n_k \leq c(M)M$ , each  $n_i \in S(c(M)M, M^{1/4-\varepsilon})$ , and*

$$\frac{A}{B} = \frac{1}{n_1} + \dots + \frac{1}{n_k}.$$

REMARK. From Lemma 3 we deduce that  $c(M) < e^{v(\varepsilon)A/B}$ , where  $v(\varepsilon)$  is some function depending only on  $\varepsilon$ . By using a “short interval” version of Lemma 1, one can prove a stronger version of Lemma 2, and possibly a stronger version of Proposition 2, which would work for all  $A/B$  with  $1/\log^{1+\varepsilon} N \ll A/B < 1$ , for any  $\varepsilon > 0$ .

Using these propositions we will now prove the Main Theorem. Let  $M$  be the least integer where

$$(3.3) \quad r \leq \sum_{N < n < M} \frac{1}{n} \leq r + \frac{1}{M}.$$

Using the fact that  $\sum_{1 \leq n \leq x} 1/n = \log x + \gamma + O(1/x)$ , it is easy to see that  $M/N = e^{r+O(1/N)}$ .

Now, from Proposition 1, we see that for  $N$  sufficiently large, there exist integers  $d_1, \dots, d_l$  with  $N < d_1 < \dots < d_l < M = e^{r+O(1/N)}N$ , such that if

$$\frac{u}{v} = \sum_{\substack{N < n < M \\ n \neq d_1, \dots, d_l}} \frac{1}{n}, \quad \gcd(u, v) = 1,$$

then all the prime power factors of  $v$  are  $\leq N^{1/5}$ . Also, from (3.2) and (3.3) we find that if we let  $A/B = r - u/v$ , where  $\gcd(A, B) = 1$ , then

$$\frac{\log \log N}{\log N} \ll \frac{A}{B} \ll \frac{\log \log N}{\log N}.$$

We observe that once  $N$  is large enough, all the prime power factors of  $B$  will be  $\leq N^{1/5}$ . We conclude from Proposition 2 with  $\varepsilon = 1/20$  that there exist integers  $n_1, \dots, n_k$  with

$$M \leq n_1 < \dots < n_k < e^{v(1/20)A/B}M,$$

where  $v(1/20)$  is some constant, and such that

$$\frac{A}{B} = \frac{1}{n_1} + \dots + \frac{1}{n_k}.$$

Thus, we have the following representation for  $r$ :

$$r = \frac{u}{v} + \frac{A}{B} = \left( \sum_{\substack{N < n < M \\ n \neq d_1, \dots, d_l}} \frac{1}{n} \right) + \frac{1}{n_1} + \dots + \frac{1}{n_k},$$

where

$$n_k < e^{v(1/20)A/B}M$$

$$= \left\{ 1 + O\left(\frac{\log \log M}{\log M}\right) \right\} M = \left\{ e^r + O_r\left(\frac{\log \log N}{\log N}\right) \right\} N.$$

This proves the first part of the Main Theorem.

To see that the  $O_r(\log \log N/\log N)$  error term is best possible, suppose that

$$r = \frac{U}{V} = \frac{1}{x_1} + \dots + \frac{1}{x_k}, \quad \text{where } \gcd(U, V) = 1.$$

Let  $Z = \max\{x_i : i = 1, \dots, k\}$ . We claim that the largest prime  $p$  dividing the  $x_i$ 's satisfies  $p < Z(1 + o_r(1))/\log Z$ . To see this, fix a prime  $p$  and suppose

$$x_1 = pm_1 < x_2 = pm_2 < \dots < x_l = pm_l$$

are all the  $x_i$ 's divisible by  $p$ . We have two cases to consider: case 1 is if  $p \mid V$ , and case 2 is when  $p \nmid V$ .

If we are in case 1, where  $p \mid V$ , then certainly  $p \leq V$ , and so  $p < Z(1 + o(1))/\log Z$ , for  $k$  sufficiently large (or  $N$  sufficiently large). If we are in case 2, where  $p \nmid V$ , then  $p \nmid Y$  either, where  $Y$  is given by

$$\frac{W}{Y} = \frac{1}{x_1} + \dots + \frac{1}{x_l} = \frac{1}{p} \left( \frac{1}{m_1} + \dots + \frac{1}{m_l} \right), \quad \gcd(W, Y) = 1.$$

Thus,  $p$  divides

$$\begin{aligned} \text{lcm}\{m_1, \dots, m_l\} \left\{ \frac{1}{m_1} + \dots + \frac{1}{m_l} \right\} \\ \leq \text{lcm}\{2, 3, \dots, m_l\} \left\{ 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{m_l} \right\} = e^{m_l(1+o(1))}, \end{aligned}$$

and so  $p < e^{m_l(1+o(1))}$ . From this we deduce that

$$Z \geq pm_l > p \log p(1 + o(1));$$

or in other words,

$$p < \frac{Z}{\log Z}(1 + o(1)).$$

Making use of this bound on  $p$  we have

$$\begin{aligned} (3.4) \quad r &= \sum_{j=1}^k \frac{1}{x_j} \leq \sum_{\substack{N < n < cN \\ p \mid n \Rightarrow p < cN(1+o(1))/\log(cN)}} \frac{1}{n} \\ &= \left( \sum_{N < n < cN} \frac{1}{n} \right) - \left( \sum_{\substack{N < mp < cN \\ p > cN(1+o(1))/\log(cN)}} \frac{1}{mp} \right). \end{aligned}$$

The first of this last pair of sums can be estimated using the well known estimate  $\sum_{x \leq n \leq y} 1/n = (\log x)/y + O(1/x)$ , which gives

$$(3.5) \quad \sum_{N < n < cN} \frac{1}{n} = \log c + O\left(\frac{1}{N}\right).$$

To estimate the second of the last pair of sums in (3.4), we will need the following lemma, which is proved at the end of this section.

LEMMA 4. For  $c > 1$  and  $\alpha > 0$  we have

$$\begin{aligned} \sum_{\substack{N < mp^\alpha < cN \\ p^\alpha > N/\log^\alpha N, p \text{ prime}}} \frac{1}{mp^\alpha} &= \sum_{\substack{N < mp < cN \\ p > N/\log^\alpha N, p \text{ prime}}} \frac{1}{mp} + O_c\left(\frac{1}{\log N}\right) \\ &= \frac{\alpha(\log c)(\log \log N)}{\log N} + O_c\left(\frac{1}{\log N}\right). \end{aligned}$$

Combining this lemma with (3.4) and (3.5), we have

$$r \leq \log c - (\log c + o(1)) \frac{\log \log N}{\log N}.$$

Solving for  $c$  we find that

$$c \geq e^r \left(1 + \frac{(r + o(1)) \log \log N}{\log N}\right).$$

*Proof of Lemma 4.* Using the fact that  $\sum_{1 \leq j \leq n} 1/j = \log n + \gamma + O(1/n)$ , together with the estimate

$$\sum_{\substack{p \leq n \\ p \text{ prime}}} \frac{1}{p} = \log \log n + \kappa + o(1/\log n),$$

where  $\kappa$  is some constant, we have the following chain of inequalities:

$$\begin{aligned} &\sum_{\substack{N < mp \leq cN \\ p > N/\log^\alpha N, p \text{ prime}}} \frac{1}{mp} \\ &= \sum_{N/\log^\alpha N < p \leq cN} \frac{1}{p} \sum_{N/p < m \leq cN/p} \frac{1}{m} \\ &= \sum_{N/\log^\alpha N < p \leq cN} \frac{1}{p} \left\{ \log\left(\frac{cN}{p}\right) - \log\left(\frac{N}{p}\right) + O\left(\frac{p}{N}\right) \right\} \\ &= \sum_{N/\log^\alpha N < p \leq cN} \frac{1}{p} \left\{ \log c + O\left(\frac{p}{N}\right) \right\} \\ &= \log c \sum_{N/\log^\alpha N < p \leq cN} \frac{1}{p} + O\left(\frac{\pi(cN)}{N}\right) \end{aligned}$$

$$\begin{aligned}
&= \log c \left\{ \log \log cN - \log \log \left( \frac{N}{\log^\alpha N} \right) + o\left( \frac{1}{\log N} \right) \right\} + O_c \left( \frac{1}{\log N} \right) \\
&= \frac{\alpha(\log c)(\log \log N)}{\log N} + O_c \left( \frac{1}{\log N} \right),
\end{aligned}$$

as claimed. The error incurred by replacing the sum over primes to a sum over prime powers will be  $O_c(1/\log N)$ .

**IV. Proof of Proposition 1.** Let  $p_1 < \dots < p_h$  be all the primes in  $[N^{1/5}, N/\log^{10} N)$ . Define

$$\begin{aligned}
S := (N, cN) \cap \mathbb{Z}, \quad S_{h+1} := S \setminus (\{mp : p \text{ prime}, p > N/\log^{10} N\} \\
\cup \{mp^a : p \text{ prime}, a \geq 2, p^a \geq N^{1/5}\}),
\end{aligned}$$

and let

$$\frac{u_{h+1}}{v_{h+1}} = \sum_{n \in S_{h+1}} \frac{1}{n}, \quad \text{where } \gcd(u_{h+1}, v_{h+1}) = 1.$$

Notice that  $v_{h+1}$  has no prime divisor  $\geq N/\log^{10} N$ ; moreover,  $v_{h+1}$  has no prime power factors  $\geq N/\log^{10} N$ , for  $N$  sufficiently large, since the only prime power divisors of elements of  $S$  that are  $\geq N^{1/5}$  are primes. We also have

$$(4.1) \quad \sum_{n \in S \setminus S_{h+1}} \frac{1}{n} = \sum_{\substack{N \leq mp \leq cN \\ p \geq N/\log^{10} N}} \frac{1}{mp} + O \left( \sum_{\substack{mp \leq cN \\ p \geq N^{1/5}, a \geq 2 \\ p \text{ prime}}} \frac{1}{mp^a} \right).$$

The first of these last two sums can be estimated using Lemma 4, which gives

$$\sum_{\substack{N \leq mp \leq cN \\ p \geq N/\log^{10} N}} \frac{1}{mp^a} = \frac{(10 \log c + o(1)) \log \log N}{\log N},$$

and the second can be estimated using Lemma 2, which gives

$$\sum_{\substack{mp^a \leq cN \\ p \geq N^{1/5}, a \geq 2 \\ p \text{ prime}}} \frac{1}{mp^a} = O \left( \frac{\log(cN)}{N^{1/10}} \right).$$

Combining the last two displayed equations with (4.1), we deduce that

$$\sum_{n \in S \setminus S_{h+1}} \frac{1}{n} = \frac{(10 \log c + o(1)) \log \log N}{\log N}.$$

Starting with the prime  $p_h$  we will successively construct subsets of  $S_{h+1}$ ,

$$S_h \supseteq S_{h-1} \supseteq S_{h-2} \supseteq \dots \supseteq S_1,$$

where if

$$\frac{u_i}{v_i} = \sum_{n \in S_i} \frac{1}{n}, \quad \gcd(u_i, v_i) = 1,$$

then all the prime factors of  $v_i$  are smaller than  $p_i$  for all  $i = 1, \dots, h$ ; moreover, we will construct these sets in such a way that

$$\sum_{n \in S_{i+1} \setminus S_i} \frac{1}{n} \ll \frac{1}{p_i \log N} \quad \text{for } i = 1, \dots, h.$$

If we can accomplish this, then if we let  $\{d_1, \dots, d_l\} = S \setminus S_1$ , we will have

$$\begin{aligned} \frac{1}{d_1} + \dots + \frac{1}{d_l} &= \sum_{n \in S \setminus S_{h+1}} \frac{1}{n} + \sum_{j=2}^{h+1} \sum_{n \in S_j \setminus S_{j-1}} \frac{1}{n} \\ &= \sum_{n \in S \setminus S_{h+1}} \frac{1}{n} + O\left(\sum_{j=1}^h \frac{1}{p_j \log N}\right) \ll \frac{\log \log N}{\log N}, \end{aligned}$$

and

$$\frac{1}{d_1} + \dots + \frac{1}{d_l} \geq \sum_{n \in S \setminus S_{h+1}} \frac{1}{n} = \frac{(10 \log c + o(1)) \log \log N}{\log N}.$$

Thus, (3.2) will be satisfied. We will also have

$$\sum_{\substack{n \in S \\ n \neq d_1, \dots, d_l}} \frac{1}{n} = \frac{u_1}{v_1},$$

where all of the prime factors of  $v_1$  are smaller than  $N^{1/5}$ ; moreover, all the prime power factors of  $v_1$  will be smaller than  $N^{1/5}$ , since the only prime powers  $\geq N^{1/5}$  that can divide elements of  $S$  are primes. Thus, (3.1) will be satisfied, and so if we can construct these sets  $S_i$ , Proposition 1 will be proved.

Suppose, for proof by induction, we have constructed the sets  $S_j$  where  $2 \leq j \leq h+1$ . Then all the prime factors of  $v_i$  are  $\leq p_{i-1}$ . If  $p_{i-1} \nmid v_i$ , we just let  $S_{i-1} := S_i$ , and then all the prime factors of  $v_{i-1}$  are smaller than  $p_{i-1}$ .

If  $p_{i-1} \mid v_i$ , then  $p_{i-1} \parallel v_i$ , since the only prime power factors of elements of  $S$  that are  $\geq N^{1/5}$  are primes. We will use Proposition 2 to construct  $S_{i-1}$  as follows: Using Bertrand's Postulate, let  $q$  be the smallest prime in  $[\log N, 2 \log N]$ , and set  $M = N/(qp_{i-1}) > (\log^9 N)/2$ . Let

$$B = \text{lcm}\{n \leq M^{1/5}\} > \text{lcm}\{n \leq (\log N)^{9/5}/2^{1/5}\} > 2cp_{i-1}M$$

(which will be true for  $M$  sufficiently large), and let  $A$  be the largest integer

$\leq c'B/2$  where

$$c' = \sum_{\substack{M < n < cM \\ n \in S(cM, M^{1/5})}} \frac{1}{n} = \varrho(5) \log c + O\left(\frac{1}{\log M}\right)$$

(which follows from Lemma 3) and

$$A \equiv qBu_i(v_i/p_{i-1})^{-1} \pmod{p_{i-1}}$$

(note:  $p_{i-1} \parallel v_i$ ). Since  $B > 2cp_{i-1}M$ , and since  $A \in [c'B/2 - p_{i-1}, c'B/2]$ , we have

$$2\frac{A}{B} \leq c' < 2\frac{A}{B} + \frac{2p_{i-1}}{B} < 2\frac{A}{B} + \frac{1}{cM},$$

for  $N$  is sufficiently large. From Proposition 2, there exist  $n_1, \dots, n_k$  with  $M < n_1 < \dots < n_k < cM$  where each  $n_i \in S(cM, M^{1/5})$  and

$$\frac{1}{n_1} + \dots + \frac{1}{n_k} = \frac{A}{B}.$$

Now, we claim that we can let

$$S_{i-1} = S_i \setminus T_i,$$

where

$$T_i = \{qp_{i-1}n_j : 1 \leq j \leq k\}.$$

Notice that the elements of  $T_i$  all lie in  $[N, cN]$ , and have largest prime divisor equal to  $p_{i-1}$ , which is their only prime power divisor  $\geq N^{1/5}$ . Thus,  $T_i \subset S_{h+1}$ . Also,  $T_i \cap T_j = \emptyset$  if  $i \neq j$ , since the largest prime divisors of elements of  $T_i$  and  $T_j$  are  $p_{i-1}$  and  $p_{j-1}$ , respectively. Thus,

$$T_i \cap S_i = T_i \cap (S_{h+1} \setminus (T_{h+1} \cup \dots \cup T_{i+1})) = T_i \cap S_{h+1} = T_i,$$

which implies  $T_i \subseteq S_i$  (in fact,  $T_i \subset S_i$ ).

If we let  $S_{i-1}$  be defined in this way, then, since  $p_{i-1} \mid v_i$ , we have

$$\frac{u_{i-1}}{v_{i-1}} = \frac{u_i}{v_i} - \frac{A}{qp_{i-1}B} = \frac{qBu_i - Av_i/p_{i-1}}{v_iqB}.$$

Thus,

$$v_iqBu_{i-1} = v_{i-1}(qBu_i - Av_i/p_{i-1}).$$

Since  $qBu_i - Av_i/p_{i-1} \equiv 0 \equiv v_i \pmod{p_{i-1}}$ , and since  $\gcd(u_{i-1}, v_{i-1}) = 1$ , we must have  $v_{i-1} \mid v_iqB/p_{i-1}$ . Now  $v_i/p_{i-1}$  is not divisible by  $p_{i-1}$ , since, as we mentioned earlier,  $p_{i-1} \parallel v_i$ , and so  $v_i/p_{i-1}$  has all its prime power divisors  $< p_{i-1}$ ; also,  $qB$  is not divisible by  $p_{i-1}$ , since  $B = \text{lcm}\{2, 3, \dots, M^{1/5}\}$ , and  $M^{1/5}$  is less than  $N^{1/5} < p_{i-1}$ , and since  $q < M$ . So, all the prime divisors of  $v_{i-1}$  are  $< p_{i-1}$ . We also have

$$\sum_{n \in S_i \setminus S_{i-1}} \frac{1}{n} < \sum_{M < n < cM} \frac{1}{qp_{i-1}n} \ll \frac{\log c}{p_{i-1}q} \ll \frac{1}{p_{i-1} \log N},$$

and so  $S_{i-1}$  satisfies all the requisite properties. We conclude that all the sets  $S_j$ ,  $j = 1, \dots, h + 1$ , can be constructed, and so Proposition 1 follows.

**V. Proof of Proposition 2.** Let

$$P := \text{lcm}(1, 2, \dots, [M^{1/4-\varepsilon}]) = e^{M^{1/4-\varepsilon}(1+o(1))},$$

where this last equality follows from the Prime Number Theorem. Let  $M \leq y_1 < \dots < y_t \leq c(M)M$  be all the divisors of  $P$  lying in  $[M, c(M)M]$ ; that is, all the integers in  $S(c(M)M, M^{1/4-\varepsilon})$  in the interval  $[M, c(M)M]$ . If  $Y \mid P$ , we have the following identity:

$$\frac{1}{P} \sum_{h=-P/2}^{P/2-1} e\left(\frac{Xh}{Y}\right) = \begin{cases} 1 & \text{if } Y \mid X, \\ 0 & \text{if } Y \nmid X, \end{cases}$$

where  $e(u) = e^{2\pi i u}$ . Thus, if  $B \mid P$ , one can deduce that

$$\begin{aligned} \#\{\{n_1, \dots, n_k\} \subseteq \{y_1, \dots, y_t\}, k \text{ variable} : 1/n_1 + \dots + 1/n_k = A/B\} \\ \geq \frac{1}{P} \sum_{h=-P/2}^{P/2-1} e\left(\frac{-Ah}{B}\right) \prod_{j=1}^t \left\{1 + e\left(\frac{h}{y_j}\right)\right\} - 2. \end{aligned}$$

(The reason for subtracting 2 in the above equation is that when  $A/B = 1$ , the exponential sum also counts the extraneous representations  $1/n_1 + \dots + 1/n_k = 0$  and 2.)

Let

$$\begin{aligned} (5.1) \quad F(h) &:= \prod_{j=1}^t \left\{1 + e\left(\frac{h}{y_j}\right)\right\} \\ &= e\left(\frac{h}{2} \left\{\frac{1}{y_1} + \dots + \frac{1}{y_t}\right\}\right) \left(2^t \prod_{j=1}^t \cos(\pi h/y_j)\right). \end{aligned}$$

Upon substituting this into our equation above this gives

$$\begin{aligned} (5.2) \quad \#\{\{n_1, \dots, n_k\} \subseteq \{y_1, \dots, y_t\}, k \text{ variable} : 1/n_1 + \dots + 1/n_k = A/B\} \\ \geq \frac{1}{P} \left( \sum_{h=-P/2}^{P/2-1} e(-Ah/B) F(h) \right) - 2. \end{aligned}$$

We will now try to find a lower bound for (5.2). To do this we will show

$$(5.3) \quad |F(h)| < \frac{2^t}{2P} \quad \text{for } M/2 < |h| \leq P/2,$$

and

$$(5.4) \quad \sum_{1 \leq h \leq M/2} e(-Ah/B) F(h) + e(Ah/B) F(-h) > 0,$$

from which we deduce

$$\sum_{0 \leq |h| \leq M/2} e(-Ah/B)F(h) > 2^t.$$

From this, (5.2), and (5.3), it then follows that

$$\begin{aligned} \#\{\{n_1, \dots, n_k\} \subseteq \{y_1, \dots, y_t\}, k \text{ variable} : 1/n_1 + \dots + 1/n_k = A/B\} \\ > \frac{2^{t-1}}{P} - 2 = 2^{t-O(M^{1/4-\epsilon})}, \end{aligned}$$

which is exponential in  $t$  since

$$t \gg_\epsilon M \frac{A}{B} \gg \frac{M \log \log \log M}{\log M}.$$

To establish (5.4), we first observe from (5.1) that

$$\begin{aligned} (5.5) \quad \text{Arg}\{e(-Ah/B)F(h)\} \\ = \frac{-2\pi Ah}{B} + \pi h \left\{ \frac{1}{y_1} + \dots + \frac{1}{y_t} \right\} + \text{Arg} \left\{ \prod_{j=1}^t \cos(\pi h/y_j) \right\}. \end{aligned}$$

Using the fact that

$$\frac{1}{y_1} + \dots + \frac{1}{y_t} = 2\frac{A}{B} + \delta,$$

where

$$0 \leq \delta \leq \frac{1}{c(M)M},$$

together with the fact that each  $y_j$  is  $\geq M$ , we have

$$(5.6) \quad \left| \frac{-2\pi Ah}{B} + \pi h \left\{ \frac{1}{y_1} + \dots + \frac{1}{y_t} \right\} \right| = \pi \delta |h| < \frac{\pi |h|}{M} \leq \frac{\pi}{2},$$

whenever

$$|h| \leq M/2.$$

Also for such  $h$ , we observe that

$$\cos(\pi h/y_j) \geq \cos(\pi/2) = 0 \quad \text{for } j = 1, \dots, t,$$

since  $y_j \geq M$  for all  $j$ . Hence,

$$\text{Arg} \left\{ \prod_{j=1}^t \cos(\pi h/y_j) \right\} = 0.$$

Using this, together with (5.5) and (5.6), we find that

$$|\text{Arg}\{e(-Ah/B)F(h)\}| < \pi/2, \quad \text{whenever } |h| < M/2.$$

Thus, for such  $h$  we have

$$e(-Ah/B)F(h) + e(Ah/B)F(-h) > 0,$$

and so (5.4) follows.

In order to establish (5.3), we will need the following lemma, which will be proved in the next section of the paper:

LEMMA 5. *Suppose  $0 < \varepsilon < 1/8$ . Let  $y_1 < \dots < y_t$  be all the integers in  $[M, (1 + 1/\log M)M]$  where each  $y_i \in S((1 + 1/\log M)M, M^{1/4-\varepsilon})$ . Then for  $M$  sufficiently large and  $h$  real, either:*

1. *There are  $\gg M^{3/4}$   $y_i$ 's which do not divide any integer in  $I := (h - M^{3/4}, h + M^{3/4})$ , or*
2. *There is an integer in this interval which is divisible by  $P := \text{lcm}\{p^a \leq M^{1/4-\varepsilon} : p \text{ prime}\}$ .*

From this lemma, it follows that if

$$M/2 \leq |h| \leq P/2,$$

and if

$$Z(c_1) = \#\{y_j, j = 1, \dots, t : \|h/y_j\| > c_1/M^{1/4}\},$$

where  $\|u\|$  denotes the distance to the nearest integer from  $u$ , then for some constants  $c_1, c_2 > 0$  we will have for all  $M$  sufficiently large,

$$Z(c_1) > c_2 M^{3/4}.$$

For these integers  $y_j$  counted by  $Z(c_1)$ , we will have

$$\begin{aligned} |\cos(\pi h/y_j)| &= |\cos(\|\pi h/y_j\|)| < |\cos(\pi c_1/M^{1/4})| \\ &= 1 - \frac{1}{2} \cdot \frac{\pi^2 c_1^2}{M^{1/2}} + O\left(\frac{1}{M}\right). \end{aligned}$$

From this and (5.1) it follows that for such  $h$ ,

$$|F(h)| < 2^t \left(1 - \frac{1}{2} \cdot \frac{\pi^2 c_1^2}{M^{1/2}} + O\left(\frac{1}{M}\right)\right)^{Z(c_1)} \ll 2^t e^{-\pi^2 c_2 c_1^2 M^{1/4}/2} = o\left(\frac{2^t}{P}\right).$$

This establishes (5.3) and thus proves the proposition.

**VI. Proof of Lemma 5.** For each integer  $n$  satisfying

$$(6.1) \quad M^{3/4} \log^3 M < n < 2M^{3/4} \log^3 M, \quad n \in S(2M^{3/4} \log^3 M, M^{1/4-\varepsilon}),$$

define

$$M(n) := \{y_j : y_j = nq, \text{ where } \omega(q) \leq 3\}.$$

We claim that  $\text{lcm } M(n) = P$  for all such  $n$ . We will show below that the truth of this claim implies that either:

A. There is an  $n$  satisfying (6.1) such that every integer of  $M(n)$  divides a single integer in  $I$ , which together with the assumption  $\text{lcm } M(n) = P$  gives us case 2 in the claim of our lemma, or

B. For each  $n$  satisfying (6.1), there is an integer  $y_{\alpha(n)} \in M(n)$  which does not divide any integer in  $(h - M^{3/4}, h + M^{3/4})$ .

We will assume that case B is true and show that it implies case 1 in the claim of our lemma (and thus if we can show that  $\text{lcm } M(n) = P$  and that either A or B is true, we may conclude that either case 1 or case 2 in our lemma is true):

The first thing to notice is that from (2.1) we know that there are at least  $c_\varepsilon M^{3/4} \log^3 M$  integers  $n$  satisfying (6.1). If all of the  $y_{\alpha(n)}$ 's as indicated in case B were distinct, then there would be at least  $c_\varepsilon M^{3/4} \log^3 M$   $y_j$ 's not dividing any integer in  $(h - M^{3/4}, h + M^{3/4})$ , which is the first possibility claimed by our lemma; however, it is not necessarily the case that the  $y_{\alpha(n)}$ 's are distinct. To overcome this difficulty, we will now show that no  $y_i$  can lie in too many of the sets  $M(n)$ : Let

$$D(M) := \max_{y_i} \#\{n : n \text{ satisfies (6.1) and } y_i \in M(n)\} \\ \leq \max_{y_i} \#\{q : q \mid y_i, \omega(q) \leq 3\} = o(\log^3 M).$$

From this we have

$$\#\{y_{\alpha(n)} : n \text{ satisfies (6.1)}\} \\ \geq \frac{\psi(2M^{3/4} \log^3 M, M^{1/4-\varepsilon}) - \psi(M^{3/4} \log^3 M, M^{1/4-\varepsilon})}{D(M)} \gg M^{3/4}.$$

Thus, there are  $\gg M^{3/4}$   $y_j$ 's which do not divide any integer in  $(h - M^{3/4}, h + M^{3/4})$ , which covers case 1 claimed by our lemma.

We will now show that if  $\text{lcm } M(n) = P$  for all  $n$  satisfying (6.1), then either case A or case B above must be true. So, let us assume that  $\text{lcm } M(n) = P$  for all  $n$  satisfying (6.1). If case B is true, then we are done. So, let us assume that case B is false. Then there is an  $n$  satisfying (6.1) such that each member of  $M(n)$  divides an integer in  $I$ . Since each such member is divisible by  $n \geq M^{3/4} \log^3 M$ , which is greater than the length of  $I$ , all such members must divide the same integer in  $I$ . Thus, case A is true.

To finish the proof of our lemma, we now show that  $\text{lcm } M(n) = P$  for all  $n$  satisfying (6.1). Fix an  $n$  satisfying (6.1) and let  $p^a \leq M^{1/4-\varepsilon}$  be the largest power of the prime  $p$  that is  $\leq M^{1/4-\varepsilon}$ . Let  $p^e$  be the exact power of  $p$  which divides  $n$ . Thus,  $e \leq a$ . We will show there exists a  $y_j \in M(n)$  with

$$y_j = np^{a-e} l_1 l_2, \quad \text{where } l_1 \text{ and } l_2 \text{ are primes with } \gcd(l_1 l_2, n) = 1,$$

which will imply that  $y_j$  is divisible by  $p^a$ , and thus  $p^a \mid \text{lcm } M(n)$ . Such a  $y_j$  exists if we can just find primes  $l_1, l_2 \leq M^{1/4-\varepsilon}$  which satisfy

$$(6.2) \quad \sqrt{\frac{M}{np^{a-e}}} \leq l_1 < l_2 \leq \sqrt{\left(1 + \frac{1}{\log M}\right) \frac{M}{np^{a-e}}}, \quad \gcd(l_1 l_2, n) = 1.$$

To see that it is possible to find  $l_1$  and  $l_2$  we first observe that the lower limit of the interval in (6.2) is

$$\sqrt{\frac{M}{np^{a-e}}} \gg \sqrt{\frac{M}{(M^{3/4} \log^3 M) M^{1/4-\varepsilon}}} = \frac{M^{\varepsilon/2}}{\log^{3/2} M},$$

and the length of the interval is the multiple  $\sqrt{1+1/\log M} - 1 \gg 1/\log M$  of this lower limit. From the Prime Number Theorem, there are  $\gg M^{\varepsilon/2}/(\varepsilon \log^{7/2} M)$  primes in this interval, and so for  $M$  sufficiently large there must be two of them  $l_1 < l_2$  which do not divide  $n < 2M^{3/4} \log^3 M$ . These two primes therefore satisfy (6.2). To see that  $l_1, l_2 < M^{1/4-\varepsilon}$ , we observe that the upper limit of the interval in (6.2) satisfies

$$\sqrt{\left(1 + \frac{1}{\log M}\right) \frac{M}{np^{a-e}}} < \sqrt{\frac{2M}{n}} \leq \sqrt{\frac{2M}{M^{3/4} \log^3 M}} = \frac{\sqrt{2} M^{1/8}}{\log^{3/2} M} < M^{1/4-\varepsilon},$$

for  $M$  sufficiently large and  $0 < \varepsilon < 1/8$ . Thus, we can find  $l_1$  and  $l_2$  as claimed, and so our lemma is proved.

**Acknowledgements.** I would like to thank Drs. Andrew Granville and Carl Pomerance for their comments and suggestions; Greg Martin, for the enlightening conversations I had with him by email and in person; and finally, Christian Elsholtz, for his many corrections and suggestions for this paper, as well as for the many interesting conversations on unit fractions.

### References

- [1] N. G. de Bruijn, *On the number of positive integers  $\leq x$  and free of prime factors  $> y$* , Nederl. Akad. Wetensch. Proc. Ser. A 54 (1951), 50–60.
- [2] E. Croot, *On some questions of Erdős and Graham*, *Mathematika*, to appear.
- [3] P. Erdős and R. L. Graham, *Old and New Problems and Results in Combinatorial Number Theory*, Monograph. Enseign. Math. 28, Univ. Genève, 1980, 33–34.
- [4] R. K. Guy, *Unsolved Problems in Number Theory*, 2nd ed., Springer, 1994, 158–166.
- [5] G. Martin, *Dense Egyptian fractions*, *Trans. Amer. Math. Soc.* 351 (1999), 3641–3657.
- [6] —, *Denser Egyptian fractions*, *Acta Arith.* 95 (2000), 231–260.

Department of Mathematics  
University of California at Berkeley  
Berkeley, CA 94720, U.S.A.  
E-mail: ecroot@math.berkeley.edu

*Received on 11.6.1999  
and in revised form on 18.9.2000*

(3622)