EGYPTIAN FRACTIONS WITH EACH DENOMINATOR HAVING THREE DISTINCT PRIME DIVISORS

Steve Butler
Department of Mathematics, Iowa State University, Ames, Iowa
butler@iastate.edu

Paul Erdős[1]
Rényi Institute, Budapest, Hungary

Ron Graham
Department of Computer Science and Engineering,
University of California, San Diego, La Jolla, California
graham@ucsd.edu

Received: 2/21/15, Accepted: 11/17/15, Published: 12/11/15

Abstract

Any natural number can be expressed as an Egyptian fraction, i.e., $\sum 1/a_i$ with $a_1 < a_2 < \cdots < a_\ell$, where each denominator is the product of three distinct primes.

1. Introduction

Egyptian fractions date back over 3500 years to the Rhind papyrus [5] (making them among the oldest mathematics still extant). They are a way to express rational numbers in a very specific form, namely

$$\frac{m}{n} = \frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_\ell}, \qquad \text{where } a_1 < a_2 < \cdots < a_\ell.$$

No one is quite certain why ancient Egyptians chose to express fractions this way (though André Weil offered the following explanation: "It is easy to explain. *They took a wrong turn!*" [1]).

The existence of Egyptian fractions for any rational number has been known since at least the time of Fibonacci (for example, the greedy algorithm will always produce a solution, though many other methods are known). However, one can

[1] Deceased.

place additional constraints on the allowable a_i and then interesting questions arise as to what is possible. In this note we will establish the following result.

Theorem 1. *Any natural number can be written as an Egyptian fraction where each denominator is the product of three distinct primes.*

A stronger version of this result for rational numbers with square-free denominators was previously mentioned by Guy [2, D11] and attributed to two of the authors (Erdős and Graham); however a proof of this result was never published! To begin correcting this situation, we give a proof of the natural number case here. We note that similar arguments could be used to show that any natural number can be written as an Egyptian fraction where each a_i is the product of ω distinct primes, for $\omega \geq 4$. We also conjecture that a similar result holds for $\omega = 2$. As an example Johnson [3] showed how 1 can be expressed as the sum of 48 unit fractions with the following denominators each of which is the product of two primes (taken from [2]):

6	21	34	46	58	77	87	115	155	215	287	391
10	22	35	51	62	82	91	119	187	221	299	689
14	26	38	55	65	85	93	123	203	247	319	731
15	33	39	57	69	86	95	133	209	265	323	901

2. Large Interval in the Sums of Products of Primes

The main work in establishing Theorem 1 will be to show that there is a large contiguous interval of integers composed of sums of products of primes. We will need the following definitions.

Definition 1. *Let p_n denote the n-th prime. Let*

$$S_n(k) = \{p_{i_1}p_{i_2}\cdots p_{i_k} : 1 \leq i_1 < i_2 < \cdots < i_k \leq n\},$$

i.e., the set of $\binom{n}{k}$ products of k distinct primes from among the first n primes. Given a set $X = \{x_1, x_2, \ldots, x_m\}$, we let

$$P(X) = \left\{ \sum_{i=1}^{m} \varepsilon_i x_i : \varepsilon_i \in \{0,1\} \right\},$$

i.e., the set of all possible subset sums involving elements of X. Finally, we let $L_n(k) = P\big(S_n(k)\big)$ and $\sigma_n(k) = \sum_{s\in S_n(k)} s$; note that $\sigma_n(k)$ is the maximal element of $L_n(k)$.

From the definition we have $L_n(k)$ is symmetric, and for some small values of n and k it can be shown by computation that there are large contiguous intervals of elements in the middle of $L_n(k)$. For example, we have the following:

- $L_8(4)$ has $\sigma_8(4) = 414849$ and contains i for $3482 \le i \le 411367$.
- $L_8(5)$ has $\sigma_8(5) = 2429223$ and contains i for $54728 \le i \le 2374495$.
- $L_8(6)$ has $\sigma_8(6) = 8130689$ and contains i for $1750114 \le i \le 6380575$.

Lemma 1. *For $n \ge 5$, $L_{n+3}(n)$ contains i for $(1/6)\sigma_{n+3}(n) \le i \le (5/6)\sigma_{n+3}(n)$.*

Before we begin the proof we state some results which will be needed.

Fact 1 (Chebyshev). *For $n \ge 1$, $p_{n+1}/p_n \le 2$.*

Theorem 2 (Olson [4]). *Let X be a set of distinct nonzero numbers modulo p, where p is a prime. If $p < (|X|^2 + 3)/4$, then $P(X)$ contains all residues modulo p.*

Fact 2. *For $n \ge 3$, the set $P\big(S_{n+2}(n+1)\big)$ contains all residues modulo p_{n+4}.*

Proof. For $3 \le n \le 13$ this was verified computationally. For $n \ge 14$ we apply Theorem 2. We note that elements in $S_{n+2}(n+1)$ are of the form $p_1 \cdots \widehat{p_i} \cdots p_{n+2}$, i.e., the product of all but one of the first $n+2$ primes. Further, these entries are all distinct ($p_1 \cdots \widehat{p_i} \cdots p_{n+2} \equiv p_1 \cdots \widehat{p_j} \cdots p_{n+2} \pmod{p_{n+4}}$ if and only if $p_j \equiv p_i$) and nonzero (since p_{n+4} does not divide any of these terms). Finally we note that $|S_{n+2}(n+1)| = n+2$. Thus we obtain all residues if $p_{n+4} < (n^2 + 4n + 7)/4$, which holds for all $n \ge 14$ (since p_n is bounded above by $n \ln n + n \ln \ln n$ (c.f. [6]) which grows much more slowly than $(n^2 + 4n + 7)/4$). $\square$

Fact 3. *For $n \ge 3$, $p_1 p_2 \cdots p_n p_{n+3} \le \sigma_{n+2}(n+1)$.*

Proof. Using the definition of $\sigma_{n+2}(n+1)$ and dividing both sides by $p_1 p_2 \cdots p_n$, this is equivalent to showing that

$$p_{n+1}p_{n+2}\left(\frac{1}{p_1} + \cdots + \frac{1}{p_n}\right) + p_{n+1} + p_{n+2} \ge p_{n+3}.$$

Since $(1/2) + (1/3) + (1/5) > 1$ we have

$$p_{n+1}p_{n+2}\left(\frac{1}{p_1} + \cdots + \frac{1}{p_n}\right) + p_{n+1} + p_{n+2} \ge p_{n+1}p_{n+2} + p_{n+2} \ge 2p_{n+2} \ge p_{n+3},$$

where at the last step we used Fact 1. $\square$

Fact 4. *For $n \ge 1$, $3\sigma_{n+2}(n+1) \le p_{n+4}\sigma_{n+3}(n)$.*

Proof. A computation verifies this for $n \le 6$, so we can assume $n \ge 7$. Consider the following (subscripts in the denominator being taken modulo $n+2$):

$$\sum_{1 \le i < j < k \le n+2} \frac{p_{n+3}p_{n+4}}{p_i p_j p_k} + \sum_{1 \le i < j \le n+2} \frac{p_{n+4}}{p_i p_j} > \sum_{1 \le i \le n+2} \left(\frac{p_{n+4}}{p_i p_{i+1}} + \frac{p_{n+4}}{p_i p_{i+2}} + \frac{p_{n+4}}{p_i p_{i+3}}\right)$$
$$> \sum_{1 \le i \le n+2} \left(\frac{1}{p_i} + \frac{1}{p_i} + \frac{1}{p_i}\right) = \sum_{1 \le i \le n+2} \frac{3}{p_i}.$$

For the first inequality, we drop our first term and most of the summands in the second, making sure that what remains is a group of distinct summands. For the second inequality, we note that each summand involves a ratio of p_{n+4} and a smaller prime, which is greater than 1. Multiplying the left and right by $p_1p_2\cdots p_{n+2}$ this becomes $p_{n+4}\sigma_{n+3}(n) > 3\sigma_{n+2}(n+1)$, as desired. $\square$

We are now ready to proceed with the proof of the main lemma.

Proof of Lemma 1. By the previously noted computation, this is true for $n = 5$ (in fact it is also true for $n = 3$ and 4).

We now proceed by induction and note that

$$L_{n+4}(n+1) = L_{n+3}(n+1) + p_{n+4}L_{n+3}(n), \tag{1}$$

i.e., the first set on the right consists of the sums not involving the largest available prime while the second set consists of the sums which do involve the largest available prime (here the "+" indicates we take all pairwise sums of the two sets). By our induction hypothesis we have that $p_{n+4}L_{n+3}(n)$ contains all multiples of p_{n+4} between $\frac{1}{6}p_{n+4}\sigma_{n+3}(n)$ and $\frac{5}{6}p_{n+4}\sigma_{n+3}(n)$.

For the first set we decompose it in the following way:

$$L_{n+3}(n+1) = P\big(S_{n+2}(n+1)\big) + P(T_n), \tag{2}$$

where $T_n = S_{n+3}(n+1) \setminus S_{n+2}(n+1)$. By Fact 2 we have that $P\big(S_{n+2}(n+1)\big)$ contains all of the residues modulo p_{n+4}.

Now let us focus on T_n, with elements $t_1 < t_2 < \cdots < t_m$. We have

$$T_n = \{p_1p_2\cdots p_np_{n+3} = t_1, t_2, \ldots, t_m = p_3p_4\cdots p_{n+3}\}.$$

We can go from t_1 to t_m by a sequence of elements of T where at each stage we replace some p_i by p_{i+1}. From Fact 1 we can conclude that the ratio of two elements in this chain is at most 2 and therefore we have $t_{i+1}/t_i \le 2$ for all i.

Because the ratios between consecutive t_i are bounded by 2 we claim

$$t_{j+1} - (t_1 + t_2 + \cdots + t_j) \le t_1 \text{ for } j \ge 1.$$

To see this we use the ratio bound to observe that

$$2t_1 + t_2 + \cdots + t_j \ge t_1 + (1/2)\big(t_2 + \cdots + t_{j+1}\big).$$

Multiplying through by 2 and rearranging then gives us the desired inequality.

We now claim that the largest gap in $P(T_n)$ is t_1. We show this by an induction on i for $P\big(\{t_1,\ldots,t_i\}\big)$. First note that for $i = 1$ we have $P\big(\{t_1\}\big) = \{0, t_1\}$, which has a gap of size t_1. Now suppose that the result holds for $P\big(\{t_1,\ldots,t_i\}\big)$ and consider

$$P\big(\{t_1,\ldots,t_i,t_{i+1}\}\big) = P\big(\{t_1,\ldots,t_i\}\big) \cup \big(t_{i+1} + P\big(\{t_1,\ldots,t_i\}\big)\big).$$

In other words $P\big(\{t_1, \ldots, t_i, t_{i+1}\}\big)$ consists of the union of $P\big(\{t_1, \ldots, t_i\}\big)$ and a translation of $P\big(\{t_1, \ldots, t_i\}\big)$ by t_{i+1}. By our induction hypothesis we know that each copy of $P\big(\{t_1, \ldots, t_i\}\big)$ has gaps bounded by at most t_1. So if there is a larger gap it must come *between* the two copies. In particular, if the two copies overlap, then we are done. Otherwise the gap between the two copies is $t_{i+1}-(t_1+t_2+\cdots+t_i)$ (minimal element in the second minus the maximal element in the first), and by the preceding claim this is at most t_1 and we are again done.

We can now rewrite (2) in the following way:

$$L_{n+3}(n+1) = \bigcup_{v_i \in P(T_n)} \big(v_i + P\big(S_{n+2}(n+1)\big)\big).$$

In particular, we have that L_{n+3} is the union of shifted copies of $P\big(S_{n+2}(n+1)\big)$.

We know by Fact 3 that $t_1 < \sigma_{n+2}(n+1)$. By Fact 2 we know that each copy of $P\big(S_{n+2}(n+1)\big)$ contains all of the residues modulo p_{n+4}.

Let $u_1, u_2, \ldots$ be the elements of $L_{n+3}(n+1)$ congruent to $r_0 \pmod{p_{n+4}}$ for some fixed r_0. Then the preceding observations allow us to conclude the following:

- $u_1 \le \sigma_{n+2}(n+1)$, since there is a copy of $P\big(S_{n+2}(n+1)\big)$ shifted by 0 in $L_{n+3}(n+1)$ and $P\big(S_{n+2}(n+1)\big)$ has all residues modulo p_{n+4}.
- $u_{i+1} - u_i \le 2\sigma_{n+2}(n+1)$, since the length of $P\big(S_{n+2}(n+1)\big)$ is more than the distance between any two consecutive values used for translation (i.e., this distance is at most t_1), every copy other than the last must intersect with the succeeding copy. Therefore in any interval of length twice the width, i.e., $2\sigma_{n+2}(n+1)$, there must be at least one full copy of $P\big(S_{n+2}(n+1)\big)$, and so also an element u_j.

Now we finally return to (1). We proceed by placing many copies of $p_{n+4}L_{n+3}(n)$ (where the middle consists of a long run of terms which differ by p_{n+4}), i.e., each element of $L_{n+3}(n+1)$ acts as a translate.

We claim that we have every element of $L_{n+4}(n+1)$ which is congruent to $r_0 \pmod{p_{n+4}}$ in the interval between $(1/6)\sigma_{n+4}(n+1)$ and $(5/6)\sigma_{n+4}(n+1)$. Since this will hold for any arbitrary r_0, we can finally conclude that $L_{n+4}(n+1)$ contains i for $(1/6)\sigma_{n+4}(n+1) \le i \le (5/6)\sigma_{n+4}(n+1)$, which will finish the proof.

We focus on the translations by u_i (forcing everything to be congruent to $r_0 \pmod{p_{n+4}}$), and show that the run of consecutive multiples of p_{n+4} will connect together and start soon enough to satisfy what we need.

To see that they connect (i.e., one run will start before the previous run finishes) we need to have for $i \ge 1$,

$$u_i + \frac{5}{6}p_{n+4}\sigma_{n+3}(n) \ge u_{i+1} + \frac{1}{6}p_{n+4}\sigma_{n+3}(n),$$

or equivalently,

$$u_{i+1} - u_i \leq \frac{2}{3} p_{n+4} \sigma_{n+3}(n).$$

We previously noted that $u_{i+1} - u_i \leq 2\sigma_{n+2}(n+1)$ and so this last relationship will hold if

$$2\sigma_{n+3}(n+1) \leq \frac{2}{3} p_{n+4} \sigma_{n+3}(n).$$

But this is precisely Fact 4. Therefore we have that these runs connect.

To see that they begin soon enough we need to have

$$\frac{1}{6} p_{n+4} \sigma_{n+3}(n) \leq \frac{1}{6} \sigma_{n+4}(n+1).$$

This is true since the left hand side can be found by summing a subset of the terms which sum to give the right hand side. Therefore we have that we will begin soon enough. (By symmetry if we start soon enough, we will also end late enough.)

We now have the needed elements modulo r_0 in the middle of $L_{n+4}(n+1)$. Since, as already noted, r_0 was arbitrary this concludes the proof of the lemma. $\square$

3. Estimating the Number of Terms

By showing that $L_{n+3}(n)$ has a contiguous long middle interval we can now finish the proof of the main result.

Proof of Theorem 1. Let m be a natural number. If for some n

$$\frac{1}{6} L_{n+3}(n) \leq p_1 p_2 \cdots p_{n+3} m \leq \frac{5}{6} L_{n+3}(n),$$

then by Lemma 1 we can conclude that

$$p_1 p_2 \cdots p_{n+3} m = \sum_S p_{i_1} p_{i_2} \cdots p_{i_n}$$

where the sum on the right hand side is over a collection of distinct sets S having the form $S = \{i_1, \ldots, i_n\} \subseteq \{1, \ldots, n+3\}$. Now dividing both sides by $p_1 p_2 \cdots p_{n+3}$ we get our desired expression of m as an Egyptian fraction where each denominator is the product of three distinct primes.

So it remains to understand the behavior of

$$\frac{L_{n+3}(n)}{p_1 p_2 \cdots p_{n+3}} = \sum_{1 \leq i < j < k \leq n+3} \frac{1}{p_i p_j p_k}.$$

But we now note that

$$\sum_{1 \leq i \leq n} \frac{1}{p_i} \sim \log \log n$$

(see [6]), and so

$$\Big(\sum_{1\le i\le n}\frac{1}{p_i}\Big)^3 \sim \big(\log\log n\big)^3.$$

But we also have that

$$\Big(\sum_{1\le i\le n}\frac{1}{p_i}\Big)^3 = \sum_{1\le i\le n}\frac{1}{p_i^3} + 2\sum_{i\ne j}\frac{1}{p_i^2 p_j} + 6\sum_{1\le i<j<k\le n+3}\frac{1}{p_ip_jp_k}.$$

The first term on the right is bounded by some constant c_1, the second term will have growth bounded by $c_2 \log\log n$ for some constant c_2, and so we can conclude that the third term drives the growth. In particular we have that the sum will go to infinity, but do so slowly and so that for any m we will find some suitable value of n so

$$\frac{1}{6}\sum_{1\le i<j<k\le n+3}\frac{1}{p_ip_jp_k} \le m \le \frac{5}{6}\sum_{1\le i<j<k\le n+3}\frac{1}{p_ip_jp_k},$$

establishing the result. $\square$

We note the proof also helps give a bound for the number of terms that will be needed. In particular, for a large m, we choose n so that

$$\frac{5}{36}\big(\log\log n\big)^3 \approx m \qquad \text{or} \qquad n \approx \exp\big(\exp(\sqrt[3]{36m/5})\big),$$

i.e., so that m falls into the appropriate interval to be able to express its location. Once n is chosen, we will then have $\binom{n+3}{n} \sim n^3/6$ possible terms to work with and in the worst case scenario we will need all of them so that we will only need at most

$$\frac{1}{6}\exp\big(3\exp(\sqrt[3]{36m/5})\big)$$

terms to express m as a sum of unit fractions whose denominators are products of three primes. Even for $m = 1$ we are unlikely to ever find the actual decomposition as it will possibly involve over one hundred million unit fractions!

We also note that the proof shows that we can write many rational numbers as Egyptian fractions with denominators the product of three primes. In particular, we have that if q is a rational number with $q' = p_1p_2\cdots p_{n+3}q$ a natural number with $\frac{1}{6}\sigma_{n+3}(n) \le q' \le \frac{5}{6}\sigma_{n+3}(n)$ (or more generally as long as $q' \in L_{n+3}(n)$), then we can also write q' as an Egyptian fraction where denominators are the products of three distinct primes.

For example we have that

$$\frac{1}{7}\cdot 2\cdot 3\cdot 5\cdot 7\cdot 11\cdot 13 = 4290 \in L(6,3),$$

so we can write $1/7$ as an Egyptian fraction whose denominators are the products of three primes. This can be done with 12 unit fractions with the following denominators:

$$30 \quad 42 \quad 66 \quad 70 \quad 78 \quad 105 \quad 110 \quad 154 \quad 165 \quad 195 \quad 273 \quad 286$$

The approach we have given here will not be enough to establish the stronger result that every rational number of the form m/n, where n is square-free, can be expressed as an Egyptian fraction whose denominators are the product of three primes. The problem is that the structure of $L_{n+3}(n)$ is not well understood at the ends of the interval. So the main challenge becomes in recovering very small rational numbers.[2]

Understanding the structure of $L_n(k)$ is related to the following problem: For each k, determine the *largest* number, $a(k)$, which cannot be written as the sum of distinct numbers each having k distinct prime divisors. For example, we have $a(1) = 6$, $a(2) = 23$, $a(3) = 299$, $a(4) = 3439$, $a(5) = 51637$, and $a(6) = 894211$. The large contiguous interval in the center of $L_n(k)$ will always start *after* $a(k)$ and so understanding this behavior gives some indication on what is happening on the end. (Indeed, one can think of $a(k) + 1$ as marking the start of the infinite contiguous interval in $L_\infty(k)$.)

Another variation that can be considered is to not use the small primes. In particular, suppose we want to only use the primes $p_r, p_{r+1}, \ldots$ for the prime divisors of the denominators. The same arguments carry through, we only need to verify the initial conditions. As an example, let

$$S'_n(k) = \big\{p_{i_1}p_{i_2}\cdots p_{i_k} : 2 \le i_1 < i_2 < \cdots < i_k \le n\big\},$$

and $L'_n(k) = P\big(S_n(k)\big)$ and $\sigma'_n(k) = \sum_{s\in S'_n(k)} s$, i.e., these are the analogous definitions of $S_n(k)$ and $L_n(k)$ except we do not use the prime 2. We then have the following:

- $L'_7(4)$ has $\sigma'_7(4) = 340419$ and contains i for $20877 \le i \le 319542$.

This is more than sufficient to begin the induction and establish the following result.

Theorem 3. *Any natural number can be written as an Egyptian fraction where each denominator is the product of three distinct* odd *primes.*

We look forward to seeing further progress in the area of Egyptian fractions.

---

[2]One of the authors believes that all rational numbers can be expressed in this form, another author has doubts that every rational number can be expressed in this form, and the third author, already having looked in *The BOOK* at the answer, remains silent on this issue.

References

[1] R. Graham, Paul Erdős and Egyptian fractions, in Erdős Centennial, L. Lovász, I. Ruzsa, V. T. Sós and D. Palvolgyi eds., Bolyai Society Mathematical Studies 25, Springer-Verlag, Heidelberg (2013), 289–309.

[2] R. K. Guy, *Unsolved Problems in Number Theory, Third Edition*, Springer-Verlag, New York, 2004.

[3] A. W. Johnson, Letter to the Editor, *Crux Mathematicorum* **4** (1978), 190.

[4] J. E. Olson, An addition theorem modulo p, *J. Combin. Theory* **5** (1968), 45–52.

[5] G. Robins and C. Shute, *The Rhind Mathematical Papyrus: An Ancient Egyptian text*, Dover, New York, 1987.

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