

Important: It has been pointed out to the authors that the bounds on $\#\mathfrak{E}_N$ given in this paper were already proved in [3] (the upper bound in a stronger form).

We plan to update further this article at a later stage.

COUNTING EGYPTIAN FRACTIONS

SANDRO BETTIN, LOIČ GRENIÉ, GIUSEPPE MOLTENI, AND CARLO SANNA

ABSTRACT. For any integer $N \geq 1$, let \mathfrak{E}_N be the set of all Egyptian fractions employing denominators less than or equal to N . We give upper and lower bounds for the cardinality of \mathfrak{E}_N , proving that

$$\frac{N}{\log N} \prod_{j=3}^k \log_j N < \log(\#\mathfrak{E}_N) < 0.421 N,$$

for any fixed integer $k \geq 3$ and every sufficiently large N , where $\log_j x$ denotes the j -th iterated logarithm of x .

1. INTRODUCTION

Every positive rational number a/b can be written in the form of an *Egyptian fraction*, that is, as a sum of distinct unit fractions: $a/b = 1/n_1 + \cdots + 1/n_r$ with $n_1, \dots, n_r \in \mathbb{N}$ distinct. Several properties of these representations have been investigated. For example, it is known that all rationals $a/b \in (0, 1)$ are representable using only denominators which are $O(b(\log b)^{1+\varepsilon})$ [11, 12] (see also [9]) or that $O(\sqrt{\log b})$ different denominators are always sufficient [10]. It is also well understood which integers can be represented using denominators up to a bound x [4], and Martin [5, 6] showed that any rational can be represented as a “dense Egyptian fraction”. In this paper we take a different direction, and study the cardinality of the set of rational numbers representable using denominators up to N ,

$$\mathfrak{E}_N := \left\{ \sum_{n=1}^N \frac{t_n}{n} : t_1, \dots, t_N \in \{0, 1\} \right\}, \quad N \in \mathbb{N},$$

as $N \rightarrow +\infty$.

Another motivation for studying the cardinality of \mathfrak{E}_N comes from the recent work of three of the authors [2] (see also [1]), where the question of how well a real number τ can be approximated by sums of the form $\sum_{n=1}^N s_n/n$, where $s_1, \dots, s_N \in \{-1, +1\}$, is studied. Precisely, let

$$\mathfrak{S}_N := \left\{ \sum_{n=1}^N \frac{s_n}{n} : s_1, \dots, s_N \in \{-1, +1\} \right\} \quad \text{and} \quad \mathfrak{m}_N(\tau) := \min \{ |\tau - \sigma| : \sigma \in \mathfrak{S}_N \},$$

for every positive integer N . Note that \mathfrak{S}_N and \mathfrak{E}_N have the same cardinality, since $x \mapsto 2x - \sum_{n=1}^N 1/n$ is a bijection $\mathfrak{E}_N \rightarrow \mathfrak{S}_N$.

It has been proved that $\mathfrak{m}_N(\tau) < \exp\left(-\left(\frac{1}{\log 4} - \varepsilon\right)(\log N)^2\right)$ for every $\tau \in \mathbb{R}$, $\varepsilon > 0$, and for all sufficiently large positive integers N , depending on τ and ε [2, Theorem 1.1]; and that for any $f: \mathbb{N} \rightarrow \mathbb{R}^+$, there exists $\tau_f \in \mathbb{R}$ such that $\mathfrak{m}_N(\tau_f) < f(N)$ for infinitely many N [1, Proposition 5.9]. On the other hand, it is possible to obtain lower bounds for $\mathfrak{m}_N(\tau)$ holding for almost all τ by giving upper bounds for the cardinality of \mathfrak{E}_N . Indeed, defining

$$\alpha := \limsup_{N \rightarrow +\infty} \frac{\log(\#\mathfrak{E}_N)}{N},$$

2010 *Mathematics Subject Classification*. Primary: 11D68, Secondary: 11B99.
Key words and phrases. Egyptian fractions.

by a Borel-Cantelli argument one obtains the following lower bound.

Lemma 1.1. *Fix $\varepsilon > 0$. For almost all $\tau \in \mathbb{R}$ we have*

$$\mathfrak{m}_N(\tau) > \exp(-(\alpha + \varepsilon)N)$$

for all sufficiently large N , depending on τ and ε .

In [2, Proposition 2.7] we proved that $\alpha < 0.6649$ with a simple argument, improving upon the trivial $\alpha \leq \log 2 = 0.69314\dots$ In this paper, we improve this bound even further, thus producing a better lower bound for $\mathfrak{m}_N(\tau)$.

Theorem 1.2. *We have $\alpha < 0.421$, that is, $\log(\#\mathfrak{E}_N) < 0.421 N$ for all N large enough.*

In the opposite direction, it is not difficult to show that $\log(\#\mathfrak{E}_N) \gg N/\log N$. In the following theorem we show that one can slightly improve over this lower bound.

Theorem 1.3. *For every integer $k \geq 3$ we have*

$$(1) \quad \log(\#\mathfrak{E}_N) > \frac{N}{\log N} \prod_{j=3}^k \log_j N$$

for all sufficiently large N , depending on k .

Theorem 1.3 is proved by showing that the set of integers N for which the quotient $\#\mathfrak{E}_N/\#\mathfrak{E}_{N-1}$ attains its maximum value 2 is quite large. With some more effort it is possible to give an explicit sequence of positive integers $(N_k)_{k \geq 1}$ such that (1) holds for every $N \geq N_k$.

Upper and lower bounds in Theorems 1.2 and 1.3 have different orders and it is not clear whether one of them seizes the true behavior of the sequence $\#\mathfrak{E}_N$. Only a few of these numbers can be computed, since the algorithms devised for this purpose have an exponential behavior (in time or in memory). The ones which are known are in Table 1 and partially appear as sequence A072207 of [8]. A graph of $\log(\#\mathfrak{E}_N)/N$ and of $\log(\#\mathfrak{E}_N)/(N/\log N)$ is in Figure 1. As the range of N for which $\#\mathfrak{E}_N$ is known is very small, the data do not make it clear whether $\log(\#\mathfrak{E}_N)/N$ might converge to 0 or to any other real number.

N	\mathfrak{E}_N	N	\mathfrak{E}_N	N	\mathfrak{E}_N	N	\mathfrak{E}_N
1	2	12	1856	23	896512	34	224129024
2	4	13	3712	24	936832	35	231010304
3	8	14	7424	25	1873664	36	237031424
4	16	15	9664	26	3747328	37	474062848
5	32	16	19328	27	7494656	38	948125696
6	52	17	38656	28	7771136	39	1896251392
7	104	18	59264	29	15542272	40	1928593408
8	208	19	118528	30	15886336	41	3857186816
9	416	20	126976	31	31772672	42	3925999616
10	832	21	224128	32	63545344	43	7851999232
11	1664	22	448256	33	112064512		

TABLE 1. The first values of $\#\mathfrak{E}_N$.

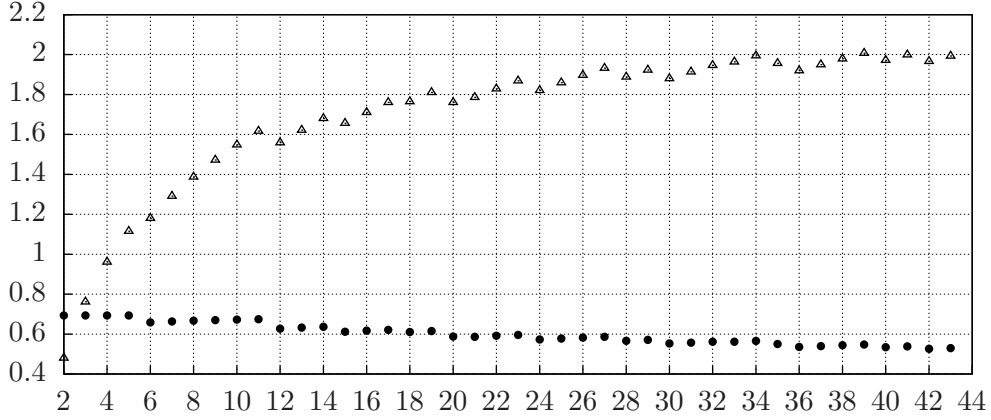


FIGURE 1. Graph of $\log(\#\mathfrak{E}_N)/N$ (dots) and of $\log(\#\mathfrak{E}_N)/(N/\log N)$ (triangles) for $N \leq 43$.

Acknowledgements. S. Bettin is member of the INdAM group GNAMPA. L. Grenié, G. Molteni and C. Sanna are members of the INdAM group GNSAGA. C. Sanna is supported by a postdoctoral fellowship of INdAM. The extensive computations needed for this paper have been performed on the UNITECH INDACO computing platform of the Università di Milano and on the computing cluster of the Université de Bordeaux. The authors warmly thank Alessio Alessi and Karim Belabas for their operative assistance for these computations.

Notation. We employ the Landau–Bachmann “Big Oh” notation O as well as the associated Vinogradov symbols \ll and \gg . We reserve the letter p for prime numbers. We put $\log_1 x := \log x$ and $\log_{k+1} x := \log(\log_k x)$ for every integer $k \geq 1$ and every sufficiently large x .

2. PROOF OF LEMMA 1.1

The definition of α implies that we have an upper bound $\#\mathfrak{S}_N \leq e^{(\alpha+\varepsilon/2)N}$, for all large enough N . The claim follows by the Borel–Cantelli lemma. We have

$$\begin{aligned} \mathcal{E} &:= \{\tau \in \mathbb{R} : \mathfrak{m}_N(\tau) \leq e^{-(\alpha+\varepsilon)N} \text{ for infinitely many } N\} \\ &= \bigcap_{M=1}^{\infty} \bigcup_{N \geq M} \{\tau \in \mathbb{R} : \mathfrak{m}_N(\tau) \leq e^{-(\alpha+\varepsilon)N}\}. \end{aligned}$$

Hence, the Lebesgue measure of \mathcal{E} is estimated by

$$\text{meas}(\mathcal{E}) \leq \inf_M \sum_{N \geq M} 2e^{-(\alpha+\varepsilon)N} \#\mathfrak{S}_N \leq \inf_M \sum_{N \geq M} 2e^{-\varepsilon N/2} = \inf_M \frac{2e^{-\varepsilon M/2}}{1 - e^{-\varepsilon/2}} = 0.$$

This implies that, for almost every τ , the lower bound $\mathfrak{m}_N(\tau) > e^{-(\alpha+\varepsilon)N}$ holds for all sufficiently large N .

3. PROOF OF THEOREM 1.2

For each prime number p and for every positive integer n , let $\nu_p(n)$ denote the p -adic valuation of n . We begin with the following easy lemma.

Lemma 3.1. *For each prime number p , let μ_p be a positive integer, and suppose that $\mu_p = 1$ for all but finitely many primes p . Then, the natural density of the set*

$$D := \{n \in \mathbb{N} : \mu_p \mid \nu_p(n) \text{ for every prime } p\}$$

is equal to

$$\delta := \prod_p \frac{1 - p^{-1}}{1 - p^{-\mu_p}},$$

where we note that only finitely many prime numbers contribute to the product. The formula can also be written as

$$\delta = \left(\sum_{d|M} \frac{1}{d} \right)^{-1} = M \left(\sum_{d|M} d \right)^{-1},$$

where $M := \prod_p p^{\mu_p - 1}$.

Proof. This is a standard argument, we reproduce it here for completeness. Let $\delta_D(n) := 1$ when $n \in D$ and $\delta_D(n) := 0$ otherwise. Note that δ_D is a multiplicative function, so that

$$\begin{aligned} F(s) &:= \sum_{n=1}^{\infty} \frac{\delta_D(n)}{n^s} = \prod_p \left(1 + \sum_{b=1}^{\infty} \frac{\delta_D(p^b)}{p^{bs}} \right) = \prod_p \left(1 + \sum_{b=1}^{\infty} \frac{1}{p^{\mu_p bs}} \right) \\ &= \prod_p \left(1 - \frac{1}{p^{\mu_p s}} \right)^{-1} = \zeta(s) \prod_p \frac{1 - p^{-s}}{1 - p^{-\mu_p s}} =: \zeta(s) H(s). \end{aligned}$$

Note that $H(s)$ is a finite Euler product (because $\mu_p = 1$ for a.e. prime), which is analytic for $\Re(s) > 0$. Set $h: \mathbb{N} \rightarrow \mathbb{R}$, with $H(s) =: \sum_{n=1}^{\infty} h(n)n^{-s}$. Then

$$\begin{aligned} \sum_{n \leq x} \delta_D(n) &= \sum_{n \leq x} \sum_{d|n} h(d) = \sum_{d \leq x} h(d) \sum_{n \leq x/d} 1 = \sum_{d \leq x} h(d) \left\lfloor \frac{x}{d} \right\rfloor = x \sum_{d \leq x} \frac{h(d)}{d} + O \left(\sum_{d \leq x} |h(d)| \right) \\ &= xH(1) + O \left(x \sum_{d > x} \frac{|h(d)|}{d} \right) + O \left(\sum_{d \leq x} |h(d)| \right) = xH(1) + O_{\varepsilon}(x^{\varepsilon}) \end{aligned}$$

and $\delta = H(1)$. (A more careful analysis shows that the error term has size $\ll (\frac{1}{\ell} \log x)^{\ell}$, uniformly in $\ell \geq 1$ and $x \geq 2$.) The alternative representation of δ as a sum of divisors of M follows immediately by the unique factorization of integers as a product of prime powers. \square

The next lemma is our key tool to provide numerical upper bounds for α .

Lemma 3.2. *Let $\mathbf{a}_1 \subset \dots \subset \mathbf{a}_{\ell} =: \mathbf{a}$ be finite nonempty sets of natural numbers. We have*

$$\alpha \leq \log 2 - \delta \sum_{i=1}^{\ell} \left(\frac{1}{\max(\mathbf{a}_i)} - \frac{1}{\max(\mathbf{a}_{i+1})} \right) \log \left(\frac{2^{\#\mathbf{a}_i}}{r_i} \right),$$

where $1/\max(\mathbf{a}_{\ell+1}) := 0$,

$$\delta := \left(\sum_{d|M} \frac{1}{d} \right)^{-1} \quad \text{with} \quad M := \text{lcm}\{a : a \in \mathbf{a}\},$$

and, for $i = 1, \dots, \ell$,

$$r_i := \#R_i \quad \text{with} \quad R_i := \left\{ \sum_{a \in \mathbf{a}_i} \frac{t_a}{a} : t_a \in \{0, 1\} \right\}.$$

Proof. For any prime p let $\mu_p := 1 + \max\{\nu_p(a) : a \in \mathbf{a}\}$ and

$$D := \{n \in \mathbb{N} : \mu_p \mid \nu_p(n) \text{ for every prime } p\}.$$

Let $k\mathbf{a} := \{ka : a \in \mathbf{a}\}$ for every integer k . Note that $k_1\mathbf{a} \cap k_2\mathbf{a} = \emptyset$ for all $k_1, k_2 \in D$ with $k_1 \neq k_2$. Indeed, suppose that $k_1 a_1 = k_2 a_2$ for some $k_1, k_2 \in D$ and $a_1, a_2 \in \mathbf{a}$. Then, for every prime number p , we have

$$\nu_p(a_1) - \nu_p(a_2) \equiv \nu_p(k_2) - \nu_p(k_1) \equiv 0 \pmod{\mu_p},$$

which in turn implies that $\nu_p(a_1) = \nu_p(a_2)$. Hence, $a_1 = a_2$ and $k_1 = k_2$.

Clearly μ_p is 1 for all but finitely many prime numbers p . Hence, Lemma 3.1 applies, and the natural density of the set D is

$$\delta = \prod_p \frac{1-p^{-1}}{1-p^{-\mu_p}} = \left(\sum_{d|M} \frac{1}{d} \right)^{-1},$$

with $M = \prod_p p^{\mu_p-1} = \text{lcm}\{a : a \in \mathbf{a}\}$. Let N be a positive integer. Since $\mathbf{a}_1, \dots, \mathbf{a}_\ell$ are contained in \mathbf{a} , we have that the sets $k\mathbf{a}_i$, with $i \in \{1, \dots, \ell\}$ and $k \in D \cap (N/\max(\mathbf{a}_{i+1}), N/\max(\mathbf{a}_i)]$ are pairwise disjoint. Let F_N be their union. Clearly, $F_N \subseteq \{1, \dots, N\}$. Also, we have

$$(2) \quad \begin{aligned} \#F_N &= \sum_{i=1}^{\ell} \# \left(D \cap \left(\frac{N}{\max(\mathbf{a}_{i+1})}, \frac{N}{\max(\mathbf{a}_i)} \right] \right) \#\mathbf{a}_i \\ &= (\delta + o(1))N \sum_{i=1}^{\ell} \left(\frac{1}{\max(\mathbf{a}_i)} - \frac{1}{\max(\mathbf{a}_{i+1})} \right) \#\mathbf{a}_i, \end{aligned}$$

as $N \rightarrow +\infty$.

We are finally ready to give an upper bound for $\#\mathfrak{E}_N$. Every element of \mathfrak{E}_N is of the form

$$\sum_{i=1}^{\ell} \sum_{k \in D \cap \left(\frac{N}{\max(\mathbf{a}_{i+1})}, \frac{N}{\max(\mathbf{a}_i)} \right]} \sum_{a \in k\mathbf{a}_i} \frac{t_a}{a} + \sum_{b \in F'_N} \frac{t_b}{b},$$

where $t_n \in \{0, 1\}$ for all $n \in \{1, \dots, N\}$, and $F'_N := \{1, \dots, N\} \setminus F_N$. Therefore, we have

$$\begin{aligned} \log \mathfrak{E}_N &\leq \sum_{i=1}^{\ell} \# \left(D \cap \left(\frac{N}{\max(\mathbf{a}_{i+1})}, \frac{N}{\max(\mathbf{a}_i)} \right] \right) \log r_i + \log(2^{N-\#F_N}) \\ &= N \left(\log 2 + (\delta + o(1)) \sum_{i=1}^{\ell} \left(\frac{1}{\max(\mathbf{a}_i)} - \frac{1}{\max(\mathbf{a}_{i+1})} \right) \log r_i \right) - \#F_N \log 2 \\ &= N \left(\log 2 - (\delta + o(1)) \sum_{i=1}^{\ell} \left(\frac{1}{\max(\mathbf{a}_i)} - \frac{1}{\max(\mathbf{a}_{i+1})} \right) \log \left(\frac{2^{\#\mathbf{a}_i}}{r_i} \right) \right), \end{aligned}$$

as $N \rightarrow +\infty$, where we used (2). Consequently,

$$\alpha \leq \log 2 - \delta \sum_{i=1}^{\ell} \left(\frac{1}{\max(\mathbf{a}_i)} - \frac{1}{\max(\mathbf{a}_{i+1})} \right) \log \left(\frac{2^{\#\mathbf{a}_i}}{r_i} \right),$$

as claimed. \square

Lemma 3.2 gives non-trivial bounds for α already when applied in the simplest case $\ell = 1$. Table 2 displays, for some choices of $\mathbf{a} = \mathbf{a}_1$, the value of the parameter $r = r_1$ (obtained numerically) and the corresponding bound $\alpha \leq \log 2 - \frac{\delta}{\max(\mathbf{a})} \log(2^{\#\mathbf{a}}/r)$.

\mathbf{a}	r	$\delta/\max(\mathbf{a})$	$\alpha \leq$
$\{1, 2, 3, 6\}$	13	1/12	0.67584390
$\{1, 2, 3, 4, 6, 12\}$	29	1/28	0.66487620
$\{1, 2, 3, 4, 5, 6, 8, 10, 12, 15\}$	302	1/45	0.66601285
$\{1, 2, 3, 4, 5, 6, 10, 12, 15, 20\}$	162	1/56	0.66022083
$\{1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 16, 20, 24, 30\}$	694	1/93	0.65915160
$\{1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 18, 20, 24, 30\}$	1061	2/195	0.65796522
$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 18, 20, 21, 24, 28, 30\}$	7757	7/780	0.65533420

TABLE 2. Upper bounds for α in the case $\ell = 1$.

Notice that it is always convenient to include 1 in \mathbf{a} since its presence does not affect the fraction $\delta/\max(\mathbf{a})$ whereas it typically increases the value of $\log(2^{\#\mathbf{a}}/r)$. The result in [2, Proposition 2.7] corresponds to the above construction with $\mathbf{a} = \{1, 2, 3, 4, 6, 12\}$. A judicious choice of a larger \mathbf{a} improves the result, but the gain in $\log(2^{\#\mathbf{a}}/r)$ is considerably tempered by the size of $\delta/\max(\mathbf{a})$, which becomes smaller and smaller. The last entry in Table 2 represents the best bound we were able to obtain with this approach.

Better results can be attained by taking $\ell > 1$. Given $1 \leq a_1 < a_2 < \dots < a_\ell$, one can take the collection $\mathbf{a}_1 \subset \mathbf{a}_2 \subset \dots \subset \mathbf{a}_\ell =: \mathbf{a}$ with $\mathbf{a}_i = \{a_1, \dots, a_i\}$ for $i = 1, \dots, \ell$. In this case the bound in Lemma 3.2 simplifies to

$$\alpha \leq \alpha_1 := \log 2 - \delta \sum_{i=1}^{\ell} \left(\frac{1}{a_i} - \frac{1}{a_{i+1}} \right) \log \left(\frac{2^i}{r_i} \right),$$

with $1/a_{\ell+1} := 0$. A first naive choice is to take $\mathbf{a}_i := \{1, \dots, i\}$ for $1 \leq i \leq \ell$. Notice that with this choice $r_i = \#\mathfrak{E}_i$. Now, for some $I \leq \ell$ let $\beta > 0$ be such that $r_i \leq e^{\beta i}$ for $I \leq i \leq \ell$. Then

$$\begin{aligned} \alpha_1 &= \log 2 - \delta \sum_{i=I}^{\ell} \left(\frac{1}{i} - \frac{1}{i+1} \right) i (\log 2 - \beta) + O(\delta \log(1 + |I|)) \\ &\leq \log 2 - \delta (\log 2 - \beta) \sum_{i=1}^{\ell} \frac{1}{i} + O(\delta \log(1 + |I|)). \end{aligned}$$

Now

$$\delta = \prod_{p \leq \ell} \frac{1 - p^{-1}}{1 - p^{-\mu_{p,\ell}}}$$

with $\mu_{p,\ell} = 1 + \max\{t : p^t \leq \ell\} = 1 + \lfloor \frac{\log \ell}{\log p} \rfloor \geq \frac{\log \ell}{\log p}$. Thus, by Merten's theorem and since

$$\sum_{p \leq \ell} \log(1 - p^{-\mu_{p,\ell}}) = \sum_{p \leq \ell} \log(1 - \ell^{-1}) \ll 1/\log(\ell + 1),$$

one obtains $\delta \sim e^{-\gamma}/\log \ell$ as $\ell \rightarrow \infty$. In particular, taking for example $I = \lfloor e^{\sqrt{\log \ell}} \rfloor$, we get

$$\alpha_1 \leq \log 2 - e^{-\gamma}(\log 2 - \beta) + o(1)$$

as $\ell \rightarrow \infty$ and so for any $\varepsilon > 0$ we obtain $\alpha_1 < \log 2 - e^{-\gamma}(\log 2 - \beta) + \varepsilon$ if ℓ is large enough. By the definition of α we have that for all $\varepsilon > 0$ one has $\beta \leq \alpha + \varepsilon$ for large enough ℓ , thus giving $\alpha_1 < \log 2 - e^{-\gamma}(\log 2 - \alpha - \varepsilon)$. Notice that due to the loss of the factor $e^{-\gamma}$ this argument fails to recover an upper bound of the type $\alpha + \varepsilon$ for α_1 , and for example it would only give $\alpha_1 \leq (1 - e^{-\gamma})\log 2 + \varepsilon = 0.3039 \dots + \varepsilon$ if $\alpha = 0$. However, it shows that by taking ℓ large enough one can surely get arbitrarily close to the upper bound $\log 2 - e^{-\gamma}(\log 2 - \alpha)$ just by performing a finite numerical computation.

An alternative and more effective approach arises from the observation that the density δ depends only on the valuations μ_p , and so it is the same for different \mathbf{a} sharing the same least common multiple. In particular, it is convenient to select the numbers in \mathbf{a} as the full collection of divisors of a given integer M . In this case the bound further simplifies to

$$(3) \quad \alpha \leq \alpha_1 \quad \text{with} \quad \alpha_1 = \delta \sum_{i=1}^{\ell} \left(\frac{1}{a_i} - \frac{1}{a_{i+1}} \right) \log r_i.$$

Indeed, if $a_1 < \dots < a_\ell$ are all the divisors of M , then we get

$$\begin{aligned} \log 2 - \delta \sum_{i=1}^{\ell} \left(\frac{1}{a_i} - \frac{1}{a_{i+1}} \right) \log(2^i) &= \left(1 - \delta \sum_{i=1}^{\ell} \left(\frac{1}{a_i} - \frac{1}{a_{i+1}} \right) i \right) \log 2 \\ &= \left(1 - \delta \sum_{i=1}^{\ell} \frac{1}{a_i} \right) \log 2 = \left(1 - \delta \sum_{d|M} \frac{1}{d} \right) \log 2 = 0. \end{aligned}$$

Numerically it seems that best results come from taking M with only “small” prime divisors (say, for example, $M = \text{lcm}\{1, \dots, m\}$, for some m , or some small variations of it) and this could be explained by the observation that with numbers of this shape an argument analogous to the above does not have a loss of $e^{-\gamma}$ (notice however that in this case r_i is not equal to $\#\mathfrak{E}_i$, so the argument is not fully recursive as in the previous case). Table 3 collects some bounds one gets in this way. Most likely further improvements could be obtained by taking larger numbers. However, the quantity of time and memory one needs to compute r_i for large i prevented us to any significant improvement on the value appearing at the bottom of Table 3.

M	$\alpha \leq$
$5040 = 2^4 \cdot 3^2 \cdot 5 \cdot 7$	0.56731289
$529200 = 2^4 \cdot 3^3 \cdot 5^2 \cdot 7^2$	0.55084208
$55440 = 2^4 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11$	0.54939823
$2116800 = 2^6 \cdot 3^3 \cdot 5^2 \cdot 7^2$	0.54552567
$4233600 = 2^7 \cdot 3^3 \cdot 5^2 \cdot 7^2$	0.54465996
$1441440 = 2^5 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11 \cdot 13$	0.53020542
$2162160 = 2^4 \cdot 3^3 \cdot 5 \cdot 7 \cdot 11 \cdot 13$	0.52779949
$4324320 = 2^5 \cdot 3^3 \cdot 5 \cdot 7 \cdot 11 \cdot 13$	0.52405384
$43243200 = 2^6 \cdot 3^3 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13$	0.51452256
$147026880 = 2^6 \cdot 3^3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17$	0.51032288
$2793510720 = 2^5 \cdot 3^3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19$	0.49944226
$13967553600 = 2^6 \cdot 3^3 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19$	0.49153796
$41902660800 = 2^6 \cdot 3^4 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19$	0.48948987

TABLE 3. Upper bounds for α using increasing collections $\mathbf{a}_j := \{a_1, \dots, a_j\}$ for $j = 1, \dots, \ell$ and $\mathbf{a} = \{d : d|M\}$ for different M .

The search of a good way to store the huge vectors containing the numbers in R_i led us to consider the following upper bound for r_i .

Lemma 3.3. *Let $\mathbf{a} := \{a_1 < a_2 < \dots < a_\ell\}$ be the full set of divisors of an integer M , let $\mathbf{a}_i := \{a_1, \dots, a_i\}$, and let R_i and r_i as in Lemma 3.2. Then*

$$r_i \leq 1 + \text{lcm}\{a_1, \dots, a_i\} \sum_{k=1}^i \frac{1}{a_k}$$

Proof. Let $L_i := \text{lcm}\{a_1, \dots, a_i\}$. Hence, $L_i/a_k \in \mathbb{Z}$ for $k = 1, \dots, i$. Consequently, $L_i R_i = \{\sum_{k=1}^i \pm L_i/a_k\} \subseteq \mathbb{Z}$. Moreover, $\max\{L_i R_i\} = -\min\{L_i R_i\} = M_i := \sum_{k=1}^i L_i/a_k$, so that $\#\{L_i R_i\} \leq 1 + 2M_i$. We can improve this bound by a factor 2 since $\pm 1 \equiv 1 \pmod{2}$, so that all numbers in $L_i R_i$ have the same parity of M_i . As a consequence, $r_i = \#R_i = \#\{L_i R_i\} \leq 1 + M_i$, as desired. \square

At this point, we are ready to explain the method that we used to prove the bound of Theorem 1.2. The main idea is to use the upper bound given by (3), but computing the exact value of r_i only for small i . Precisely, let M be a (very large) positive integer and let $a_1 < \dots < a_\ell$ be all its divisors. Also, let M' be a (small) divisor of M and let $a'_1 < \dots < a'_m$ be all its divisors. For each divisor a'_j of M' we pre-compute the value $r'_j := \left\{ \sum_{a \in \mathbf{a}'_j} \frac{t_a}{a} : t_a \in \{0, 1\} \right\}$, where $\mathbf{a}'_j := \{a'_1, \dots, a'_j\}$. Then, for each divisor a_i of M we look for the largest a'_j dividing a_i , and we estimate r_i with the minimum between what we get from Lemma 3.3 and the number $r'_j \cdot 2^{\#\mathbf{a}_i - \#\mathbf{a}'_j}$. This is a correct bound since, by the assumption on j , we have $\mathbf{a}'_j \subseteq \mathbf{a}_i$ and consequently r_i is at most r'_j multiplied by the power of two elevated to the difference of the cardinalities of \mathbf{a}_i and \mathbf{a}'_j .

Table 4 collects some results we get in this way for suitable choices of M' and M . The last entry gives our best result, which proves Theorem 1.2. Its computation needed approximatively 40 hours and 200GB.

M	M'	$\alpha \leq$
$\text{lcm}\{1, \dots, 17\}$	$\text{lcm}\{1, \dots, 10\}$	0.52408774
$\text{lcm}\{1, \dots, 31\}$	$\text{lcm}\{1, \dots, 17\}$	0.47705355
$\text{lcm}\{1, \dots, 41\}$	$\text{lcm}\{1, \dots, 17\}$	0.46441548
$\text{lcm}\{1, \dots, 53\}$	$\text{lcm}\{1, \dots, 17\}$	0.44930098
$\text{lcm}\{1, \dots, 79\}$	$\text{lcm}\{1, \dots, 19\}$	0.43238142
$\text{lcm}\{1, \dots, 89\}$	$\text{lcm}\{1, \dots, 19\}$	0.42691423
$\text{lcm}\{1, \dots, 97\}$	$\text{lcm}\{1, \dots, 23\}$	0.42447521
$\text{lcm}\{1, \dots, 97\}$	$2 \cdot 3 \cdot \text{lcm}\{1, \dots, 19\}$	0.42310594
$\text{lcm}\{1, \dots, 97\}$	$2^2 \cdot 3^2 \cdot \text{lcm}\{1, \dots, 19\}$	0.42286665
$\text{lcm}\{1, \dots, 97\}$	$2^2 \cdot 3^2 \cdot 5 \cdot \text{lcm}\{1, \dots, 19\}$	0.42099405

TABLE 4. Upper bounds for α mixing the computation of the true value of r_i for small divisors, and estimating its value with Lemma 3.3 for large divisors. The relevant parameters are M and M' .

4. PROOF OF THEOREM 1.3

Define the set

$$\mathcal{U} := \left\{ N \geq 1 : \sum_{n=1}^{N-1} \frac{w_n}{n} \neq \frac{1}{N} : \forall w_1, \dots, w_{N-1} \in \{-1, 0, +1\} \right\},$$

and let $\mathcal{U}(x) := \mathcal{U} \cap [1, x]$ for every $x \geq 1$.

Lemma 4.1. *If $N \in \mathcal{U}$ then $\#\mathfrak{E}_N = 2\#\mathfrak{E}_{N-1}$. Consequently, $\#\mathfrak{E}_N \geq 2^{\#\mathcal{U}(N)}$.*

Proof. On the one hand, $N \in \mathcal{U}$ implies that $\mathfrak{E}_{N-1} \cap (\mathfrak{E}_{N-1} + 1/N) = \emptyset$. On the other hand, $\mathfrak{E}_N = \mathfrak{E}_{N-1} \cup (\mathfrak{E}_{N-1} + 1/N)$. Hence, we have $\#\mathfrak{E}_N = \#\mathfrak{E}_{N-1} + \#(\mathfrak{E}_{N-1} + 1/N) = 2\#\mathfrak{E}_{N-1}$, as claimed. \square

In light of Lemma 4.1, to produce a lower bound for $\#\mathfrak{E}_N$ it is sufficient to give a lower bound for $\#\mathcal{U}(N)$.

Lemma 4.2. *\mathcal{U} contains 1 and all prime numbers.*

Proof. The fact that $1 \in \mathcal{U}$ follows immediately from the definition of \mathcal{U} . Furthermore, for every prime number p and for every $w_1, \dots, w_{p-1} \in \{-1, 0, +1\}$, the identity

$$\sum_{n=1}^{p-1} \frac{w_n}{n} = \frac{1}{p}$$

is impossible, since the left-hand side has nonnegative p -adic valuation (all denominators are not divisible by p) while the right-hand side has negative p -adic valuation. Hence, $p \in \mathcal{U}$. \square

For each positive integer m , let $d_m := \text{lcm}\{1, \dots, m\}$ and $g_m := d_m \sum_{j=1}^m 1/j$.

Lemma 4.3. *If $m \in \mathcal{U}$ and $p > g_m$ is a prime number, then $mp \in \mathcal{U}$.*

Proof. Suppose by contradiction that $N := mp \notin \mathcal{U}$. Hence, there exist $w_1, \dots, w_{N-1} \in \{-1, 0, +1\}$ such that

$$\sum_{n=1}^{N-1} \frac{w_n}{n} = \frac{1}{N}.$$

Consequently, splitting the sum according to whether $p|n$ and multiplying by p we obtain

$$p \sum_{\substack{n=1 \\ p \nmid n}}^{N-1} \frac{w_n}{n} + \sum_{j=1}^{m-1} \frac{w_{pj}}{j} = \frac{1}{m},$$

which in turn implies that

$$(4) \quad \sum_{j=1}^{m-1} \frac{w_{pj}}{j} - \frac{1}{m} \equiv 0 \pmod{p}.$$

Since $m \in \mathcal{U}$, the number on the left hand side is non-zero. Moreover, the absolute value of its numerator is at most g_m , which by hypothesis is strictly smaller than p . But then (4) is impossible. \square

Remark 4.1. The proofs of Lemma 4.2 and Lemma 4.4 can be easily adapted to show that if $m \in \mathcal{U}$ and $p > g_m$ is a prime number then $mp^k \in \mathcal{U}$ for every positive integer k . However, this generalization does not lead to an improvement of our final result.

Lemma 4.4. *If $m \in \mathcal{U}$ and $p > 3^m$ is a prime number, then $mp \in \mathcal{U}$.*

Proof. In light of Lemma 4.3, it is enough to prove that $g_m < 3^m$ for every positive integer m . On the one hand, from [7, p. 228] we know that $d_m = \exp(\psi(m)) \leq \exp(1.04m)$. On the other hand,

$$\sum_{j=1}^m \frac{1}{j} \leq 1 + \int_1^m \frac{dt}{t} = \log(em).$$

Hence, $g_m \leq \exp(1.04m) \log(em) < 3^m$ for every integer $m \geq 25$. A direct computation shows that the $g_m < 3^m$ holds also for $m = 1, \dots, 24$. \square

As usual, let $\pi(x)$ denote the number of prime numbers not exceeding x . The next lemma gives a recursive lower bound for $\#\mathcal{U}(x)$.

Lemma 4.5. *Let $y \geq 1$ and $x \geq 3^y$. Then*

$$\#\mathcal{U}(x) \geq \sum_{m \in \mathcal{U}(y)} \pi\left(\frac{x}{m}\right) - 2 \cdot 3^y.$$

Proof. Let us consider the natural numbers of the form mp , where p is a prime number satisfying $3^y < p \leq x/m$ and $m \in \mathcal{U}(y)$. Thanks to Lemma 4.4, we have that $mp \in \mathcal{U}(x)$. Moreover, these numbers can be written in the form mp in a unique way. Indeed, for the sake of contradiction, suppose that $mp = m'p'$ for some m', p' satisfying the same conditions as m, p , with $p \neq p'$. Then, $p' | m$, so that $3^y < p' \leq m \leq y$, which is impossible. At this point, counting the choices for m and p , we get

$$\#\mathcal{U}(x) \geq \sum_{m \in \mathcal{U}(y)} \left(\pi\left(\frac{x}{m}\right) - \pi(3^y) \right) \geq \sum_{m \in \mathcal{U}(y)} \pi\left(\frac{x}{m}\right) - \pi(3^y)y.$$

The desired claim follows by applying the inequality $\pi(3^y)y \leq 2 \cdot 3^y$, which in turn follows from the estimate $\pi(x) \leq 2x/\log x$ valid for all $x \geq 2$. \square

We need the following technical lemma. Let $e_1 := 1$ and $e_{k+1} := e^{e^k}$ for all integers $k \geq 2$.

Lemma 4.6. *For any fixed integer $k \geq 2$, we have*

$$\int_{e_k}^x \frac{1}{t \log t} \prod_{j=3}^k \log_j t \, dt \sim \prod_{j=2}^k \log_j x,$$

as $x \rightarrow +\infty$.

Proof. For $k = 2$ the empty product appearing on the left hand side is set to 1, by definition, and the claim is clear in this case. Assume $k \geq 3$. We have

$$\begin{aligned} \left(\prod_{j=2}^k \log_j x \right)' &= \left(\prod_{j=2}^k \log_j x \right) \sum_{j=2}^k \prod_{i=0}^j \frac{1}{\log_i x} \\ &= \frac{1}{x \log x} \left(\prod_{j=3}^k \log_j x \right) \left(1 + \sum_{j=3}^k \prod_{i=3}^j \frac{1}{\log_j x} \right) \sim \frac{1}{x \log x} \left(\prod_{j=3}^k \log_j x \right), \end{aligned}$$

as $x \rightarrow +\infty$. The claim then follows from de l'Hôpital's rule. \square

Lemma 4.7. *For every integer $k \geq 2$, we have*

$$\#\mathcal{U}(x) \gg_k \frac{x}{\log x} \prod_{j=3}^k \log_j x,$$

for all sufficiently large x , depending on k .

Proof. We proceed by induction on k . By Lemma 4.2 and by Chebyshev's estimate, we have

$$\#\mathcal{U}(x) > \pi(x) \gg \frac{x}{\log x},$$

for all sufficiently large x . This proves the claim for $k = 2$. Suppose $k \geq 3$ and that we have already proved the claim for $k - 1$. Put $y := \frac{1}{2} \log x$ and assume that x is sufficiently large. By Lemma 4.5 and by Chebyshev's estimate, we have

$$(5) \quad \#\mathcal{U}(x) \geq \sum_{m \in \mathcal{U}(y)} \pi\left(\frac{x}{m}\right) - 2 \cdot 3^y \gg \sum_{\substack{m \in \mathcal{U}(y) \\ m \geq 2}} \frac{x}{m \log(x/m)} > \frac{x}{\log x} \sum_{\substack{m \in \mathcal{U}(y) \\ m \geq 2}} \frac{1}{m}.$$

Furthermore, by partial summation and by the induction hypothesis, we get

$$(6) \quad \sum_{\substack{m \in \mathcal{U}(y) \\ m \geq 2}} \frac{1}{m} \gg \int_{e_{k-1}}^y \frac{\#\mathcal{U}(t)}{t^2} dt \gg_k \int_{e_{k-1}}^y \frac{1}{t \log t} \prod_{j=3}^{k-1} \log_j t dt \gg \prod_{j=2}^{k-1} \log_j y \gg \prod_{j=3}^k \log_j x,$$

for all sufficiently large x , depending on k , where we employed Lemma 4.6. Putting together (5) and (6) we obtain the desired claim. \square

Theorem 1.3 follows easily by Lemma 4.1 and 4.7.

REFERENCES

1. S. Bettin, G. Molteni, and C. Sanna, *Greedy approximations by signed harmonic sums and the Thue–Morse sequence*, <https://arxiv.org/abs/1805.00075>.
2. S. Bettin, G. Molteni, and C. Sanna, *Small values of signed harmonic sums*, C. R. Math. Acad. Sci. Paris **356** (2018), no. 11–12, 1062–1074.
3. M. N. Bleicher and P. Erdős, *The number of distinct subsums of $\sum_1^N 1/i$* , Math. Comp. **29** (1975), 29–42, Collection of articles dedicated to Derrick Henry Lehmer on the occasion of his seventieth birthday.
4. E. S. Croot, III, *On some questions of Erdős and Graham about Egyptian fractions*, Mathematika **46** (1999), no. 2, 359–372.
5. G. Martin, *Dense Egyptian fractions*, Trans. Amer. Math. Soc. **351** (1999), no. 9, 3641–3657.
6. G. Martin, *Denser Egyptian fractions*, Acta Arith. **95** (2000), no. 3, 231–260.
7. B. Rosser, *Explicit bounds for some functions of prime numbers*, Amer. J. Math. **63** (1941), 211–232.
8. N. J. A. Sloane, *The On-Line Encyclopedia of Integer Sequences*, <http://oeis.org>.
9. G. Tenenbaum and H. Yokota, *Length and denominators of Egyptian fractions. III*, J. Number Theory **35** (1990), no. 2, 150–156.
10. M. D. Vose, *Egyptian fractions*, Bull. London Math. Soc. **17** (1985), no. 1, 21–24.
11. H. Yokota, *On a conjecture of M. N. Bleicher and P. Erdős*, J. Number Theory **24** (1986), no. 1, 89–94.
12. H. Yokota, *On a problem of Bleicher and Erdős*, J. Number Theory **30** (1988), no. 2, 198–207.

DIPARTIMENTO DI MATEMATICA, UNIVERSITÀ DI GENOVA, VIA DODECANESO 35, 16146 GENOVA, ITALY
E-mail address: `bettin@dima.unige.it`

DIPARTIMENTO DI INGEGNERIA GESTIONALE, DELL'INFORMAZIONE E DELLA PRODUZIONE, UNIVERSITÀ DI
BERGAMO, VIALE MARCONI 5, 24044 DALMINE, ITALY
E-mail address: `loic.grenie@gmail.com`

DIPARTIMENTO DI MATEMATICA, UNIVERSITÀ DI MILANO, VIA SALDINI 50, 20133 MILANO, ITALY
E-mail address: `giuseppe.molteni1@unimi.it`

DIPARTIMENTO DI MATEMATICA, UNIVERSITÀ DI GENOVA, VIA DODECANESO 35, 16146 GENOVA, ITALY
E-mail address: `carlo.sanna.dev@gmail.com`