TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 351, Number 9, Pages 3641–3657
S 0002-9947(99)02327-2
Article electronically published on March 22, 1999

DENSE EGYPTIAN FRACTIONS

GREG MARTIN

Abstract. Every positive rational number has representations as Egyptian fractions (sums of reciprocals of distinct positive integers) with arbitrarily many terms and with arbitrarily large denominators. However, such representations normally use a very sparse subset of the positive integers up to the largest denominator. We show that for every positive rational there exist representations as Egyptian fractions whose largest denominator is at most N and whose denominators form a positive proportion of the integers up to N, for sufficiently large N; furthermore, the proportion is within a small factor of best possible.

1. Introduction

The ancient Egyptians wrote positive rational numbers as sums of distinct reciprocals of positive integers, or unit fractions. In 1202, Fibonacci published an algorithm (subsequently rediscovered by Sylvester in 1880, among others) for constructing such representations, which have come to be called Egyptian fractions, for any positive rational number. Since that time, number theorists have been interested in some quantitative aspects of Egyptian fraction representations. For instance, there are algorithms that improve upon the Fibonacci–Sylvester algorithm in various ways, bounding the size of the largest denominator or limiting the number of terms. Bleicher [1] has a thorough survey of, and references to, such developments.

One line of questions concentrates on the number of terms in Egyptian fraction representations. A positive rational m/n can always be expanded into an Egyptian fraction with at most m terms, for instance by the Farey series algorithm (see Golomb [3]). Erdős and Straus conjectured that, for all integers $n \geq 2$, the fraction $4/n$ can actually be written as the sum of three unit fractions rather than four. (In their conjecture, distinctness of the terms is not required.) Sierpiński [6] made the same conjecture for fractions of the form $5/n$, and mentioned that Schinzel conjectured that, for any fixed numerator m, the fraction m/n could be written as the sum of three unit fractions for sufficiently large n. In this vein, Vaughan [7] showed that almost all positive fractions with numerator m can be written as the sum of three unit fractions.

One can also investigate the behavior of the number of terms in Egyptian fractions at the other extreme. Any unit fraction can be split into two using the identity $1/n = 1/(n+1) + 1/(n(n+1))$; consequently, given a particular Egyptian fraction representation, one can repeatedly use this splitting algorithm on the term with

Received by the editors July 7, 1997.
1991 *Mathematics Subject Classification.* Primary 11D68.

largest denominator to construct such representations with arbitrarily many terms. However, this process results in a tremendously thin set of denominators in the sense that if the largest denominator is x, then the number of denominators is $\ll \log\log x$. One can try to use the splitting algorithm on the intermediate terms, but then the danger arises that the resulting unit fractions will no longer be distinct.

It is therefore interesting to ask how many terms can be used to represent a rational number as an Egyptian fraction, given a bound on the size of the largest denominator. The purpose of this paper is to demonstrate that a positive proportion of the integers up to the bound can in fact be assembled to form such a representation. We will establish the following theorem:

Theorem 1. *Let r be a positive rational number and $\eta > 0$ a real number. For any real number x that is sufficiently large in terms of r and η, there is a set $\mathcal{S}$ of positive integers not exceeding x such that $r = \sum_{n\in\mathcal{S}} 1/n$ and $|\mathcal{S}| > (C(r) - \eta)x$, where*

$$C(r) = (1 - \log 2)\Big(1 - \exp\Big(\frac{-r}{1 - \log 2}\Big)\Big).$$

That one can always find such a dense set of denominators is already surprising (though the Egyptians might not have been pleased to do their arithmetic this way!), but it turns out that the proportion given in Theorem 1 is comparable to the best possible result. Given a positive rational number r, any set $\mathcal{S}$ of positive integers not exceeding x with cardinality $\lfloor(1 - e^{-r})x\rfloor$ satisfies

$$\sum_{n\in\mathcal{S}} 1/n \ge \sum_{e^{-r}x+1<n\le x} 1/n = r + O_r(x^{-1}). \tag{1}$$

Consequently, the best possible value for $C(r)$ in Theorem 1 would be $1 - e^{-r}$. Notice that $C(r)/(1 - e^{-r}) = 1 - O(r)$ as r tends to zero, and that

$$\frac{C(r)}{1 - e^{-r}} = (1 - \log 2)\frac{1 - \exp(-r/(1 - \log 2))}{1 - e^{-r}} > 1 - \log 2 = 0.30685\ldots.$$

Thus when r is small, the value for $C(r)$ given in Theorem 1 is very nearly best possible, and in any case it is never smaller than 30% of the best possible value.

While the proof of Theorem 1 involves a rather complex notation, the idea underlying the construction is quite straightforward. One begins by subtracting from r the reciprocals of a suitable set $\mathcal{A}$ of integers not exceeding x with cardinality at least $(C(r) - \eta)x$. For a given prime p dividing the denominator of this difference, we can add back in the reciprocals of a very few members of $\mathcal{A}$ so that the factors of p in the denominator are cancelled out. If we repeat this process for all large primes, we will have expressed r as the sum of the reciprocals of almost all of the elements of $\mathcal{A}$, plus a small rational number whose denominator is only divisible by small primes. We are then able to write this small rational number as the sum of reciprocals of distinct integers much smaller than the members of $\mathcal{A}$, using the idea outlined above combined with an existing algorithm for expanding rational numbers into Egyptian fractions.

For a given prime p and rational number r whose denominator is a multiple of p, we might have to add the reciprocals of as many as $p-1$ multiples of p to eliminate the factor of p from the denominator of r, as shown in Lemmas 2 and 3. Since all of the elements of $\mathcal{A}$ are at most x in size, we must have $p(p-1) \le x$; thus we must

restrict our attention to integers that are roughly $x^{1/2}$-smooth. Fortunately, the distribution of smooth numbers has been widely studied. For instance, the density of $x^{1/2}$-smooth integers of size x is $1 - \log 2$ (see Section 3); this is the origin of such factors in the expression for $C(r)$.

Moreover, it turns out that we must consider integers that are divisible by different powers of p separately, which forces us to have at least $p-1$ multiples of p^2 in our set $\mathcal{A}$, at least $p-1$ multiples of p^3, etc. Thus we must restrict to $x^{1/2}$-smooth integers that are squarefree with respect to primes exceeding $x^{1/3}$, cubefree with respect to primes exceeding $x^{1/4}$, and so on. We shall find that the extra conditions on the multiplicity of their prime factors is easily handled. For simplicity we shall work with, roughly, $x^{1/2}$-smooth numbers that are k-free (not divisible by the kth power of any prime) and that are squarefree with respect to primes exceeding $x^{1/k}$, for some integer $k \geq 2$. Such integers satisfy all the above constraints, and we show in Lemma 8 that the multiples of each prime power among such integers are sufficiently plentiful for our argument to go through.

We define $\log_1 x = \max\{\log x, 1\}$ and $\log_j x = \log_1 (\log_{j-1} x)$ for any integer $j \geq 2$, and we write $\log^k x$ and $\log_j^k x$ as shorthand for $(\log x)^k$ and $(\log_j x)^k$ respectively. We use the notations $P(n)$ and $p(n)$ to denote the greatest and least prime factors of n respectively, and make the conventions that $P(1) = 1$ and $p(1) = \infty$. We say that a prime power p^l exactly divides an integer n, or that n is exactly divisible by p^l, if p^l divides n but p^{l+1} does not. The constants implicit in the $\ll$ and O-notations in this paper may depend on any Greek variable (α, δ, ε, η, and λ) and also on k and the rational number r or a/b where appropriate, but they will never depend on p, q, t, v, w, x, or y; this remains the case when any of these variables is adorned with primes or subscripts. When the phrase "x is sufficiently large" is used, the size of x may depend on the Greek variables and k, r, and a/b as well.

The author would like to thank Hugh Montgomery and Trevor Wooley for their guidance and support and the referee for many valuable comments. This material is based upon work supported under a National Science Foundation Graduate Research Fellowship.

2. Elementary lemmas

The following lemma is an easy consequence of the Cauchy–Davenport Theorem (see [8, Lemma 2.14], for instance), but we provide a direct proof.

Lemma 2. *Let t be a nonnegative integer, and let $x_1, \dots, x_t$ be nonzero elements of $\mathbf{Z}_p$, not necessarily distinct. Then the number of elements of $\mathbf{Z}_p$ that can be written as the sum of some subset (possibly empty) of the x_i is at least $\min\{p, t+1\}$. In particular, if $t \geq p-1$, then every element of $\mathbf{Z}_p$ can be so written.*

We remark that the conclusion of the lemma is best possible, since we could take $x_1 \equiv \cdots \equiv x_t \not\equiv 0 \pmod p$, in which case the set of elements of $\mathbf{Z}_p$ that can be written as the sum of a subset of the x_i is precisely $\{0, x_1, \dots, tx_1\}$, which has cardinality $t+1$ if $t < p$.

Proof. We use induction on t, the case $t = 0$ being trivial. Given a positive integer t and nonzero elements $x_1, \dots, x_t$ of $\mathbf{Z}_p$, let $\mathcal{S}_i$ be the set of elements of $\mathbf{Z}_p$ that can be written as the sum of some subset of $x_1, \dots, x_i$; then certainly $\mathcal{S}_i \subset \mathcal{S}_t$ and so $|\mathcal{S}_i| \leq |\mathcal{S}_t|$. By induction we may assume that $|\mathcal{S}_{t-1}| \geq \min\{p, t\}$. If $|\mathcal{S}_{t-1}| = p$, then $|\mathcal{S}_t| = p$ as well, so we may assume that $p > |\mathcal{S}_{t-1}| \geq t$.

Suppose that $|\mathcal{S}_t| < t+1$. Then we have $|\mathcal{S}_t| \le t \le |\mathcal{S}_{t-1}| \le |\mathcal{S}_t|$, and so $|\mathcal{S}_t| = |\mathcal{S}_{t-1}| = t < p$. In particular, the map $f : \mathcal{S}_{t-1} \to \mathcal{S}_t$ defined by $f(y) = y + x_t$ is actually a bijection of $\mathcal{S}_{t-1}$ onto itself. Thus

$$\sum_{y\in\mathcal{S}_{t-1}} y \equiv \sum_{y\in\mathcal{S}_{t-1}} (y+x_t) \equiv tx_t + \sum_{y\in\mathcal{S}_{t-1}} y \pmod p,$$

which implies that $tx_t \equiv 0 \pmod p$, a contradiction since x_t is nonzero $\pmod p$ and $0 < t < p$. Therefore $|\mathcal{S}_t| \ge t+1$, which establishes the lemma. □

Using this lemma, we can show that a factor of a prime p can be eliminated from the denominator of a rational number by adding the reciprocals of fewer than p integers from any prescribed set meeting certain conditions.

Lemma 3. *Let p^l be a prime power, and let N be an integer that is exactly divisible by p^l. Let c/d be a rational number with d dividing N, and let $\mathcal{S}$ be a set of integers dividing N that are all exactly divisible by p^l. If $|\mathcal{S}| \ge p-1$, then there is a subset $\mathcal{T}$ of $\mathcal{S}$ with cardinality less than p such that, if we define*

$$\frac{c'}{d'} = \frac{c}{d} + \sum_{n\in\mathcal{T}} \frac{1}{n}$$

with c'/d' in lowest terms, then d' divides N/p.

Proof. Without loss of generality we may assume that $|\mathcal{S}| = p-1$, whereupon we denote the elements of $\mathcal{S}$ by $n_1, \dots, n_{p-1}$. Let $M = \operatorname{lcm}\{d, n_1, \dots, n_{p-1}\}$, let $m = M/d$, and let $m_i = M/n_i$ for each $1 \le i \le p-1$. Note that M divides N and that each m_i is nonzero $\pmod p$. By Lemma 2, every element of $\mathbf{Z}_p$ can be written as the sum of some subset of $m_1, \dots, m_{p-1}$. In particular, we can choose distinct indices $i_1, \dots, i_t$ for some $0 \le t \le p-1$ so that

$$m_{i_1} + \cdots + m_{i_t} \equiv -cm \pmod p. \tag{2}$$

If we set $\mathcal{T} = \{n_{i_1}, \dots, n_{i_t}\}$, then

$$\frac{c}{d} + \sum_{n\in\mathcal{T}} \frac{1}{n} = \frac{cm}{M} + \sum_{j=1}^{t} \frac{m_{i_j}}{M} = \frac{(cm + m_{i_1} + \cdots + m_{i_t})/p}{M/p},$$

where the numerator is an integer by virtue of equation (2). Since M divides N, this establishes the lemma. □

We now cite an existing Egyptian fraction algorithm, which we will use near the end of the proof of Theorem 1.

Lemma 4. *Let c/d be a positive rational number with d odd. Suppose that $c/d < 1/P(d)$. Then there exists a set $\mathcal{C}$ of distinct odd positive integers, with*

$$\max\{n \in \mathcal{C}\} \ll d \prod_{p\le P(d)} p,$$

such that $c/d = \sum_{n\in\mathcal{C}} 1/n$.

Proof. Breusch has shown that any positive rational number with odd denominator can be written as the sum of reciprocals of distinct odd positive integers. When

one examines his construction [2, Lemmas 1–3], one finds that, when $c/d < 1/P(d)$, the integers involved in fact do not exceed

$$5\,\mathrm{lcm}\{d,\, 3^2 \prod_{3<p\le P(d)} p\},$$

which implies the bound given in the statement of the lemma. □

3. Smooth number sets

For real numbers $x,\, y > 1$, we recall the definition of $\mathcal{A}(x,y)$, the set of y-smooth numbers up to x:

$$\mathcal{A}(x,y) = \{n \le x : P(n) \le y\}.$$

Let $w > 1$ and $0 \le \lambda < 1$ be real numbers and $k \ge 2$ be an integer. We will need to work with the following sets of smooth numbers with various specified properties:

(3)
$$\begin{aligned}
\mathcal{A}(x,y;w,\lambda) &= \{n : \lambda x < n \le x;\ P(n) \le y;\ n \text{ is } k\text{-free};\ d^2 \mid n \Rightarrow P(d) \le w\},\\
\mathcal{A}(x,y;w,\lambda;p^l) &= \{n \in \mathcal{A}(x,y;w,\lambda) : n = mp^l \text{ with } P(m) < p\},\\
\mathcal{A}_0(x,y;w,\lambda) &= \{n \in \mathcal{A}(x,y;w,\lambda) : n \text{ is odd, not of the form } m^2+m-1\},\\
\mathcal{A}_0(x,y;w,\lambda;p^l) &= \mathcal{A}_0(x,y;w,\lambda) \cap \mathcal{A}(x,y;w,\lambda;p^l).
\end{aligned}$$

The first of these sets is the set of k-free smooth numbers that are squarefree with regard to large primes, as described in the introduction, while the second is the subset of the first consisting of those integers whose largest prime factor is p with multiplicity exactly l. The sets $\mathcal{A}(x,y;w,\lambda;p^l)$ and $\mathcal{A}_0(x,y;w,\lambda;p^l)$ do not actually depend on the parameter y as long as $y \ge p$; but we retain the parameter y for consistency of notation, and as a reminder that $\mathcal{A}(x,y;w,\lambda;p^l) \subset \mathcal{A}(x,y;w,\lambda)$ and likewise for the $\mathcal{A}_0$-sets.

We note that, for any $y' < y$, we may write $\mathcal{A}(x,y;w,\lambda)$ as the disjoint union

$$\begin{aligned}
\mathcal{A}(x,y;w,\lambda) = \mathcal{A}(x,y';w,\lambda) &\cup \Big(\bigcup_{\substack{y'<p\le y\\ p>w}} \mathcal{A}(x,y;w,\lambda;p) \Big)\\
&\cup \Big(\bigcup_{\substack{y'<p\le y\\ p\le w}} \bigcup_{l=1}^{k-1} \mathcal{A}(x,y;w,\lambda;p^l) \Big),
\end{aligned}$$

and the same is true if we replace every occurrence of $\mathcal{A}$ by $\mathcal{A}_0$.

As is customary, we let $\Psi(x,y)$ denote the cardinality of $\mathcal{A}(x,y)$. We will also use Ψ, with any list of arguments and with or without subscript, to denote the cardinality of the corresponding $\mathcal{A}$-set. It is well-known that $\Psi(x,y) \sim \rho(\log x/\log y)x$ for a certain range of x and y, where the Dickman rho-function $\rho(u)$ is defined for $u > 0$ to be the continuous solution to the differential-difference equation

(4)
$$\begin{aligned}
\rho(u) &= 1, \quad 0 < u \le 1;\\
u\rho'(u) &= -\rho(u-1), \quad u > 1.
\end{aligned}$$

The following lemma, due to Hildebrand, describes this asymptotic formula more specifically. We have not made any effort to optimize the error term or the range of y in the hypothesis, as it will suffice for our purposes as stated.

Lemma 5. *Let $x \ge 3$ and y be real numbers satisfying $\exp(\log_2^4 x) \le y \le x$. Then*

$$\Psi(x,y) = x\rho\Big(\frac{\log x}{\log y}\Big)\Big(1 + O\Big(\frac{\log_2 x}{\log y}\Big)\Big).$$

Proof. Hildebrand shows [5, Theorem 1] that, for any $\varepsilon > 0$, one has

$$\Psi(x, x^{1/u}) = x\rho(u)\Big(1 + O\Big(\frac{u \log_1 u}{\log x}\Big)\Big)$$

uniformly for $x \ge 3$ and $1 \le u \le \log x/(\log_2 x)^{5/3+\varepsilon}$. On setting $\varepsilon = 7/3$ and $u = \log x/\log y$ (so that $y = x^{1/u}$), the lemma follows immediately. □

From the definition (4) of $\rho(u)$, one can easily derive that ρ is a positive, decreasing function, and that $\rho(u) = 1 - \log u$ for $1 \le u \le 2$, so that $\rho(2) = 1 - \log 2$ in particular. We will need the following additional properties of $\rho(u)$.

Lemma 6. *Let $x \ge 3$, $v \ge 1$, and y be real numbers satisfying $v \le x$ and $\exp(\log_2^4 x) \le y \le x$. Then:*

(a) $\rho\Big(\dfrac{\log x/v}{\log y}\Big) \ll \rho\Big(\dfrac{\log x}{\log y}\Big) v^{\log_2 x/\log y}$;

(b) *for any $\varepsilon > 0$, we have* $\rho\Big(\dfrac{\log x}{\log y}\Big)^{-1} \ll x^{\varepsilon}$;

(c) *for any real number $\alpha > 1$, we have*

$$\sum_{\log x < n \le x^{1/\alpha}} n^{-\alpha} \rho\Big(\frac{\log x n^{-\alpha}}{\log y}\Big) \ll \rho\Big(\frac{\log x}{\log y}\Big)(\log x)^{-\alpha+1};$$

(d) *if* $\log v = o\Big(\dfrac{\log y}{\log_2^2 x}\Big)$, *then* $\rho\Big(\dfrac{\log x/v}{\log y}\Big) = \rho\Big(\dfrac{\log x}{\log y}\Big)\Big(1 + O\Big(\dfrac{\log v \log_2^2 x}{\log y}\Big)\Big)$.

Proof. Hildebrand shows [5, Lemma 1(vi)] that, for any $0 \le t \le u$, we have

$$\rho(u - t) \ll \rho(u)(u \log_1^2 u)^t. \tag{5}$$

Part (a) follows immediately on taking $u = \log x/\log y$ and $t = \log v/\log y$ and noting that $u \log_1^2 u \le \log x$. Part (b) follows from part (a) on taking $v = x$ and noting that

$$x^{\log_2 x/\log y} \le x^{1/\log_2^3 x} \ll x^{\varepsilon}.$$

By using part (a) again, the sum in part (c) is

$$\ll \rho\Big(\frac{\log x}{\log y}\Big) \sum_{\log x < n \le x^{1/\alpha}} n^{-\alpha} (n^{\alpha})^{\log_2 x/\log y} \ll \rho\Big(\frac{\log x}{\log y}\Big) \sum_{n > \log x} n^{-\alpha+\alpha/\log_2^3 x}. \tag{6}$$

For x sufficiently large, the exponent $-\alpha + \alpha/\log_2^3 x$ is bounded above away from -1. Therefore the right-hand side of (6) is

$$\begin{aligned} &\ll \rho\Big(\frac{\log x}{\log y}\Big)(\log x)^{-\alpha+\alpha/\log_2^3 x+1} \\ &= \rho\Big(\frac{\log x}{\log y}\Big)(\log x)^{-\alpha+1} \exp(\alpha/\log_2^2 x) \ll \rho\Big(\frac{\log x}{\log y}\Big)(\log x)^{-\alpha+1}, \end{aligned}$$

which establishes part (c).

Finally, for any real numbers $1 < s < t$, we have

$$\rho(s) - \rho(t) = -\int_s^t \rho'(u)\,du = \int_s^t \frac{\rho(u-1)}{u}\,du \\ \ll \frac{\rho(s-1)}{s}(t-s) \ll \rho(s)(s\log_1^2 s)^1\Big(\frac{t-s}{s}\Big)$$

by equation (5). Therefore

$$\rho(t) = \rho(s)(1 + O((t-s)\log_1^2 s)).$$

Letting $s = \log(x/v)/\log y$ and $t = \log x/\log y$, we see that

$$\rho\Big(\frac{\log x}{\log y}\Big) = \rho\Big(\frac{\log x/v}{\log y}\Big)\Big(1 + O\Big(\frac{\log v \log_2^2 x}{\log y}\Big)\Big);$$

and under the hypothesis that $\log v = o(\log y/\log_2^2 x)$, the $(1 + O(\cdot))$-term can be inverted and moved to the other side of the equation, which establishes part (d). □

With Lemmas 5 and 6 at our disposal, we can establish the following lemmas concerning the distributions of the sets of smooth numbers defined in equation (3).

Lemma 7. *Let $x \ge 3$, y, and w be real numbers satisfying $\exp(\log_2^4 x) \le y \le x$ and $w \ge \log x$. Then:*

(a) $\Psi(x,y;w,0) = \dfrac{x}{\zeta(k)}\rho\Big(\dfrac{\log x}{\log y}\Big)\Big(1 + O\Big(\dfrac{\log_2^3 x}{\log y}\Big)\Big)$;

(b) $\Psi_0(x,y;w,0) = \dfrac{x}{\xi(k)}\rho\Big(\dfrac{\log x}{\log y}\Big)\Big(1 + O\Big(\dfrac{\log_2^3 x}{\log y}\Big)\Big)$, *where* $\xi(k) = 2(1-2^{-k})\zeta(k)$.

Proof. From the definition (3) of $\mathcal{A}(x,y;w,\lambda)$, we have

$$x^{-1}\Psi(x,y;w,0) = x^{-1}\sum_{\substack{n\le x\\ P(n)\le y}}\Big(\sum_{\substack{d^k\mid n\\ P(d)\le w}}\mu(d)\Big)\Big(\sum_{\substack{f^2\mid n\\ p(f)>w}}\mu(f)\Big).$$

Interchanging the order of summation yields

$$\begin{aligned} x^{-1}\Psi(x,y;w,0) &= x^{-1}\sum_{\substack{d\le x^{1/k}\\ P(d)\le\min\{y,w\}}}\mu(d)\sum_{\substack{f\le\sqrt{xd^{-k}}\\ P(f)\le y\\ p(f)>w}}\mu(f)\sum_{\substack{n\le x\\ P(n)\le y\\ d^k\mid n\\ f^2\mid n}}1 \\ &= \sum_{\substack{d\le x^{1/k}\\ P(d)\le\min\{y,w\}}}\mu(d)\sum_{\substack{f\le\sqrt{xd^{-k}}\\ P(f)\le y\\ p(f)>w}}\mu(f)(x^{-1}\Psi(xd^{-k}f^{-2},y)) \\ &= \sum_{\substack{d\le x^{1/k}\\ P(d)\le\min\{y,w\}}}\mu(d)d^{-k}\sum_{\substack{f\le\sqrt{xd^{-k}}\\ P(f)\le y\\ p(f)>w}}\mu(f)f^{-2}\rho\Big(\frac{\log xd^{-k}f^{-2}}{\log y}\Big)\Big(1 + O\Big(\frac{\log_2 x}{\log y}\Big)\Big), \end{aligned}$$

where the final equality follows from Lemma 5. The primary contribution to the double sum will come from those terms with d small and $f = 1$ (notice that if

$f > 1$, then $f > w \ge \log x$), so we write it as

$$x^{-1}\Psi(x,y;w,0) = \sum_{d<\log x} \mu(d)d^{-k}\rho\Big(\frac{\log xd^{-k}}{\log y}\Big)\Big(1+O\Big(\frac{\log_2 x}{\log y}\Big)\Big) + O\bigg(\sum_{\log x\le d\le x^{1/k}} d^{-k}\rho\Big(\frac{\log xd^{-k}}{\log y}\Big) + \sum_{d=1}^{\infty} d^{-k}\sum_{\log x\le f\le\sqrt{x}} f^{-2}\rho\Big(\frac{\log xf^{-2}}{\log y}\Big)\bigg). \tag{7}$$

The two final terms can be estimated by Lemma 6(c):

$$\sum_{\log x\le d\le x^{1/k}} d^{-k}\rho\Big(\frac{\log xd^{-k}}{\log y}\Big) \ll \rho\Big(\frac{\log x}{\log y}\Big)\log^{-k+1} x,$$

$$\sum_{d=1}^{\infty} d^{-k}\sum_{\log x\le f\le\sqrt{x}} f^{-2}\rho\Big(\frac{\log xf^{-2}}{\log y}\Big) \ll \rho\Big(\frac{\log x}{\log y}\Big)\log^{-1} x\sum_{d=1}^{\infty} d^{-k} = \zeta(k)\rho\Big(\frac{\log x}{\log y}\Big)\log^{-1} x.$$

Also, for $d < \log x$, Lemma 6(d) gives us

$$\rho\Big(\frac{\log xd^{-k}}{\log y}\Big) = \rho\Big(\frac{\log x}{\log y}\Big)\Big(1+O\Big(\frac{(\log(\log^k x))\log_2^2 x}{\log y}\Big)\Big).$$

Thus equation (7) becomes

$$x^{-1}\Psi(x,y;w,0) = \rho\Big(\frac{\log x}{\log y}\Big)\bigg(\Big(1+O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big)\sum_{d<\log x}\mu(d)d^{-k} + O\big((\log x)^{-k+1}+\log^{-1} x\big)\bigg).$$

On noting that

$$\sum_{d<\log x}\mu(d)d^{-k} = \sum_{d=1}^{\infty}\mu(d)d^{-k} + O\bigg(\sum_{d\ge\log x} d^{-k}\bigg) = \zeta(k)^{-1} + O(\log^{-k+1} x),$$

part (a) of the lemma is established.

Let $\mathcal{A}_1(x,y;w,0)$ and $\mathcal{A}_2(x,y;w,0)$ denote the odd and even elements, respectively, of $\mathcal{A}(x,y;w,0)$. For a set $\mathcal{S}$ and an integer n, write $n\cdot\mathcal{S} = \{ns : s\in\mathcal{S}\}$ and note that $|n\cdot\mathcal{S}| = |\mathcal{S}|$. We see that

$$\begin{aligned} 2\cdot\mathcal{A}(x/2,y;w,0) &= \{2n : n\le x/2;\ n \text{ is } k\text{-free};\ P(n)\le y;\ d^2\mid n\Rightarrow P(d)\le w\} \\ &= \{n : n\le x;\ n \text{ is even and } k\text{-free};\ P(n)\le y;\ d^2\mid n\Rightarrow P(d)\le w\} \\ &\quad\cup\{2^k n : n\le x/2^k;\ n \text{ is odd and } k\text{-free};\ P(n)\le y;\ d^2\mid n\Rightarrow P(d)\le w\} \\ &= \mathcal{A}_2(x,y;w,0)\cup 2^k\cdot\mathcal{A}_1(x/2^k,y;w,0) \end{aligned}$$

as a disjoint union. Taking cardinalities on both sides, we see that

$$\Psi(x/2,y;w,0) = \Psi_2(x,y;w,0) + \Psi_1(x/2^k,y;w,0),$$

or equivalently

$$\Psi_1(x,y;w,0) - \Psi_1(x/2^k,y;w,0) = \Psi(x,y;w,0) - \Psi(x/2,y;w,0),$$

since $\Psi = \Psi_1 + \Psi_2$. Consequently, for any nonnegative integer m, we have

$$\begin{aligned}\Psi_1(x,y;w,0) &- \Psi_1(x/2^{(m+1)k},y;w,0)\\ &= \sum_{j=0}^{m}\big(\Psi_1(x/2^{jk},y;w,0) - \Psi_1(x/2^{(j+1)k},y;w,0)\big)\\ &= \sum_{j=0}^{m}\big(\Psi(x/2^{jk},y;w,0) - \Psi(x/2^{jk+1},y;w,0)\big).\end{aligned} \tag{8}$$

Choose $m = \lfloor \log_2 x/(k\log 2)\rfloor$, so that $2^{mk} \le \log x < 2^{(m+1)k}$. For any $0 \le l \le (m+1)k$, part (a) gives us

$$\begin{aligned}\Psi(x/2^l,y;w,0) &= \frac{x}{2^l\zeta(k)}\rho\Big(\frac{\log(x/2^l)}{\log y}\Big)\Big(1+O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big)\\ &= \frac{x}{2^l\zeta(k)}\rho\Big(\frac{\log x}{\log y}\Big)\Big(1+O\Big(\frac{\log(2^k\log x)\log_2^2 x}{\log y}\Big)\Big)\Big(1+O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big)\end{aligned}$$

by Lemma 6(d) and the choice of m. Using this in equation (8) gives us

$$\begin{aligned}\Psi_1(x,y;w,0) &= \sum_{j=0}^{m}\Big(\frac{x}{2^{jk}\zeta(k)} - \frac{x}{2^{jk+1}\zeta(k)}\Big)\rho\Big(\frac{\log x}{\log y}\Big)\Big(1+O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big)\\ &\quad + O(\Psi(x/2^{(m+1)k},y;w,0))\\ &= \frac{x}{2\zeta(k)}\rho\Big(\frac{\log x}{\log y}\Big)\Big(1+O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big)\sum_{j=0}^{m}\frac{1}{2^{jk}} + O\Big(\frac{x}{2^{(m+1)k}}\rho\Big(\frac{\log x}{\log y}\Big)\Big)\\ &= \frac{x}{2\zeta(k)}\rho\Big(\frac{\log x}{\log y}\Big)\Big(1+O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big)\big((1-2^{-k})^{-1}+O\big(\frac{1}{\log x}\big)\big)\\ &\quad + O\Big(\frac{x}{\log x}\rho\Big(\frac{\log x}{\log y}\Big)\Big)\\ &= \frac{x}{\xi(k)}\rho\Big(\frac{\log x}{\log y}\Big)\Big(1+O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big).\end{aligned}$$

Finally, the number of integers not exceeding x of the form m^2+m-1 is at most $\sqrt{x}$. Therefore,

$$\begin{aligned}\Psi_0(x,y;w,0) &= \Psi_1(x,y;w,0) + O(\sqrt{x})\\ &= \frac{x}{\xi(k)}\rho\Big(\frac{\log x}{\log y}\Big)\Big(1+O\Big(\frac{\log_2^3 x}{\log y} + x^{-1/2}\rho\Big(\frac{\log x}{\log y}\Big)^{-1}\Big)\Big).\end{aligned}$$

Since the second error term is $\ll x^{-1/2+\varepsilon}$ by Lemma 6(b), it is dominated by the first error term, and so part (b) is established. □

Lemma 8. *Let $k \ge 2$ be an integer and $0 < \varepsilon < 1/k$ and $0 < \lambda < 1$ be real numbers. Let x be a sufficiently large real number, and let y and w be real numbers and p a prime satisfying $\log x \le w \le x^{(1-\varepsilon)/k}$ and $\exp(\log_2^4 x) \le p < y \le x^{(1-\varepsilon)/2}$. Let $l < k$ be a positive integer, with $l = 1$ if $p \ge w$. Then:*

(a) $\Psi(x,y;w,\lambda) = (1-\lambda)\dfrac{x}{\zeta(k)}\rho\Big(\dfrac{\log x}{\log y}\Big)\Big(1+O\Big(\dfrac{\log_2^3 x}{\log y}\Big)\Big)$;

(b) $\Psi_0(x,y;w,\lambda) = (1-\lambda)\dfrac{x}{\xi(k)}\rho\Big(\dfrac{\log x}{\log y}\Big)\Big(1+O\Big(\dfrac{\log_2^3 x}{\log y}\Big)\Big)$;

(c) $\Psi(x,y;w,\lambda;p^l) > p$, *and if* $p \ne 2$, *then* $\Psi_0(x,y;w,\lambda;p^l) > p$;
(d) *there exists an element of* $\mathcal{A}(x,y;w,\lambda)$ *that is exactly divisible by* 2^l *(for any* $1 \le l < k$*).*

Proof. Clearly $\Psi(x,y;w,\lambda) = \Psi(x,y;w,0) - \Psi(\lambda x,y;w,0)$ and similarly for the Ψ_0, and so parts (a) and (b) follow from Lemma 7 on noting that

$$\rho\Big(\frac{\log \lambda x}{\log y}\Big) = \rho\Big(\frac{\log x}{\log y}\Big)\Big(1 + O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big)$$

by Lemma 6(d) with $v = \lambda^{-1}$.

We notice that $mp^l \in \mathcal{A}(x,y;w,\lambda;p^l)$ if and only if $m \in \mathcal{A}(xp^{-l}, p-1; w, \lambda)$, and so

$$\begin{aligned}\Psi(x,y;w,\lambda;p^l) &= \Psi(xp^{-l}, p-1; w, \lambda)\\ &= \rho\Big(\frac{\log xp^{-l}}{\log(p-1)}\Big)\frac{(1-\lambda)xp^{-l}}{\zeta(k)}\Big(1 + O\Big(\frac{\log_2^3 x}{\log(p-1)}\Big)\Big)\end{aligned}$$

by part (a). To show that this is greater than p, it suffices to show that

$$\frac{\zeta(k)p^{l+1}}{1-\lambda}\rho\Big(\frac{\log xp^{-l}}{\log(p-1)}\Big)^{-1} < x\Big(1 + O\Big(\frac{\log_2^3 x}{\log(p-1)}\Big)\Big). \tag{9}$$

But under the restrictions on p, l, y, and w, we have that $p^{l+1} < x^{1-\varepsilon}$; and by Lemma 6(b) we have that $\rho(\log xp^{-l}/\log(p-1))^{-1} \le \rho(\log x/\log(p-1))^{-1} \ll x^{\varepsilon/2}$. Therefore, the left-hand side of equation (9) is $\ll x^{1-\varepsilon/2}$. This establishes the first assertion of part (c), and the proof of the second assertion is similar.

Finally, by part (b) with x replaced by $x/2^l$, we see that $\Psi_0(x/2^l, y; w, \lambda)$ is positive for sufficiently large x. But we can choose an element m of $\mathcal{A}_0(x/2^l, y; w, \lambda)$, and then $2^l m$ is an element of $\mathcal{A}(x,y;w,\lambda)$ that is exactly divisible by 2^l, which establishes part (d). □

Lemma 9. *Under the hypotheses of Lemma 8, we have*

(a) $$\sum_{n \in \mathcal{A}(x,y;w,\lambda)} n^{-1} = \rho\Big(\frac{\log x}{\log y}\Big)\frac{\log \lambda^{-1}}{\zeta(k)}\Big(1 + O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big);$$

(b) $$\sum_{n \in \mathcal{A}_0(x,y;w,\lambda)} n^{-1} = \rho\Big(\frac{\log x}{\log y}\Big)\frac{\log \lambda^{-1}}{\xi(k)}\Big(1 + O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big);$$

(c) *if* $y = y(x)$ *is chosen so that* $\log x/\log y \le B$ *for some constant* B *depending only on* k*, then for any positive real number* α*, there exists a real number* $0 < \lambda < 1$*, bounded away from zero uniformly in* x*, such that when* x *is sufficiently large, we have*

$$0 < \alpha - \sum_{n \in \mathcal{A}_0(x,y;w,\lambda)} n^{-1} < x^{-1}\exp(D\alpha) \tag{10}$$

for some constant D *depending only on* k*.*

When we apply part (c) we will need only that the middle expression in equation (10) is positive and $\ll x^{-1}$, but we have been explicit about the upper bound for two reasons. First, since we will use part (c) to choose a value of λ, it is important that the implicit constant not depend on λ. Second, we will apply part (c) with a very small value of α, and one that is specified only up to an error bounded by

a function of x. From the form of the upper bound in equation (10), we will be assured that this will not inflate the implicit constant as x grows.

Proof. We have

(11)
$$\begin{aligned}\zeta(k)\sum_{n\in\mathcal{A}(x,y;w,\lambda)} n^{-1} &= \zeta(k)\int_{\lambda x}^{x} t^{-1}\,d\Psi(t,y;w,0)\\ &= \zeta(k)\,t^{-1}\Psi(t,y;w,0)\big|_{\lambda x}^{x} + \zeta(k)\int_{\lambda x}^{x}\Psi(t,y;w,0)\,\frac{dt}{t^2}\\ &= \Big(\rho\Big(\frac{\log x}{\log y}\Big) - \rho\Big(\frac{\log \lambda x}{\log y}\Big) + \int_{\lambda x}^{x}\rho\Big(\frac{\log t}{\log y}\Big)\frac{dt}{t}\Big)\Big(1+O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big)\end{aligned}$$

by Lemma 7(a). However, since $\log\lambda^{-1} = O(1) = o(\log y/\log_2^2 x)$, we can apply Lemma 6(d) to see that

$$\rho\Big(\frac{\log t}{\log y}\Big) = \rho\Big(\frac{\log x}{\log y}\Big)\Big(1+O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big)$$

uniformly for $\lambda x \le t \le x$. Therefore equation (11) becomes

$$\begin{aligned}\zeta(k)\sum_{n\in\mathcal{A}(x,y;w,\lambda)} n^{-1} &= \rho\Big(\frac{\log x}{\log y}\Big)\Big(O\Big(\frac{\log_2^3 x}{\log y}\Big) + \int_{\lambda x}^{x}\frac{dt}{t}\Big)\Big(1+O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big)\\ &= \rho\Big(\frac{\log x}{\log y}\Big)\log\lambda^{-1}\Big(1+O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big).\end{aligned}$$

This establishes part (a), and the proof of part (b) is exactly similar.

As for part (c), note that the function $\alpha - \sum_{n\in\mathcal{A}_0(x,y;w,t)} n^{-1}$ is an increasing function of t, with jump discontinuities of size not exceeding $(tx)^{-1}$, and which takes negative values if x is sufficiently large. Therefore we can choose λ such that

$$0 < \alpha - \sum_{n\in\mathcal{A}_0(x,y;w,\lambda)} n^{-1} \le (\lambda x)^{-1}. \tag{12}$$

For this value of λ, part (b) tells us that

$$\alpha > \sum_{n\in\mathcal{A}_0(x,y;w,\lambda)} n^{-1} = \frac{\log\lambda^{-1}}{\xi(k)}\rho\Big(\frac{\log x}{\log y}\Big)\Big(1+O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big),$$

which implies that

$$\lambda^{-1} < \exp\left(\alpha\frac{\xi(k)}{\rho(B)}\Big(1+O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big)\right).$$

This shows that λ is bounded away from zero uniformly in x; and combining this bound with equation (12) and writing $D = 2\xi(k)/\rho(B)$ establishes equation (10) for sufficiently large x. □

4. Construction of dense Egyptian fractions

We now proceed with the proof of Theorem 1. Let $r = a/b$ be a positive rational number and η a positive real number. For the first stage of the proof, let $0 < \varepsilon < 1/2$

and $0 < \delta < r$ be real numbers and $k \ge 2$ an integer that is large enough to ensure that b is k-free; we will later constrain these parameters in terms of r and η. Set

$$\lambda = \exp\Big(-\frac{(r-\delta)\zeta(k)}{\rho(2/(1-\varepsilon))}\Big). \tag{13}$$

Let $x \ge 3$ be a sufficiently large real number, set $y = x^{(1-\varepsilon)/2}$ and $w = x^{(1-\varepsilon)/k}$, and for $z > 2$ define

$$D(z) = D(z;w) = \Big(\prod_{\substack{p<z\\ p\le w}} p^{k-1}\Big)\Big(\prod_{w<p<z} p\Big),$$
$$D_0(z) = 2^{-k+1}D(z),$$

so that $D_0(z)$ is the odd part of $D(z)$. (We understand that the second product in the definition of $D(z)$ is 1 if $z \le w$.) Notice that if p is a prime, then p does not divide $D(p)$.

Let $p_1 > p_2 > \cdots > p_R$ be the primes in $(w, y]$, and for later consistency of notation, let p_0 be the smallest prime exceeding y. Notice that all elements of $\mathcal{A}(x, y; w, 0)$ divide $D(p_0)$, as does b as long as $P(b) \le w$; we henceforth assume that x is large enough to ensure this. Define

$$\frac{a_0}{b_0} = r - \sum_{n\in\mathcal{A}(x,y;w,\lambda)} \frac{1}{n}$$

with a_0/b_0 in lowest terms. Notice that b_0 divides $D(p_0)$ and that

$$\frac{a_0}{b_0} = \delta + O\Big(\frac{\log_2^3 x}{\log y}\Big) \tag{14}$$

by Lemma 9(a) and the choice (13) of λ.

We now recursively construct a sequence of fractions a_1/b_1, a_2/b_2, … , a_R/b_R with the following two properties. First, each a_i/b_i is obtained from the previous a_{i-1}/b_{i-1} by adding the reciprocals of a few elements of $\mathcal{A}(x, y; w, \lambda)$, specifically elements of $\mathcal{A}(x, y; w, \lambda; p_i)$; so when we have constructed these fractions, we will have written r as the sum of a_R/b_R and the reciprocals of almost all of the elements of $\mathcal{A}(x, y; w, \lambda)$. Second, all of the primes dividing each denominator b_i will be less than p_i, so the denominators are becoming gradually smoother.

Formally, we construct the fractions a_i/b_i as follows. Given a_{i-1}/b_{i-1}, where $1 \le i \le R$, we apply Lemma 3 with $p = p_i$, $l = 1$, $N = D(p_{i-1})$, $c/d = a_{i-1}/b_{i-1}$, and $\mathcal{S} = \mathcal{A}(x, y; w, \lambda; p_i)$. This gives us a subset of $\mathcal{A}(x, y; w, \lambda; p_i)$, which we call $\mathcal{B}_i$, with cardinality less than p_i such that, if we define

$$\frac{a_i}{b_i} = \frac{a_{i-1}}{b_{i-1}} + \sum_{n\in\mathcal{B}_i} \frac{1}{n}$$

with a_i/b_i in lowest terms, then b_i divides $D(p_{i-1})/p_i = D(p_i)$. We are justified in applying Lemma 3 with these parameters since, by Lemma 8(c), the size of $\mathcal{A}(x, y; w, \lambda; p_i)$ is at least p_i.

Now set $y' = \exp(\log_2^4 x)$, and let $q_1 > q_2 > \cdots > q_S$ be the primes in $[y', w]$; and for later consistency of notation, set $q_0 = p_R$ and $a_{0,k}/b_{0,k} = a_R/b_R$. We recursively construct another sequence of fractions $a_{1,1}/b_{1,1}$, $a_{1,2}/b_{1,2}$, … , $a_{1,k}/b_{1,k}$, $a_{2,1}/b_{2,1}$, … , $a_{S,k}/b_{S,k}$ with similar properties to the first sequence, the differences coming from the fact that the primes q_i may potentially divide the denominators we are

working with to the $(k-1)$st power. First, each $a_{i,1}/b_{i,1}$ is equal to $a_{i-1,k}/b_{i-1,k}$, and each $a_{i,j}/b_{i,j}$ is obtained from the previous $a_{i,j-1}/b_{i,j-1}$ by adding the reciprocals of a few elements of $\mathcal{A}(x,y;w,\lambda)$, specifically elements of $\mathcal{A}(x,y;w,\lambda;q_i^{k-j+1})$; again, we will have written r as the sum of $a_{S,k}/b_{S,k}$ and the reciprocals of almost all of the elements of $\mathcal{A}(x,y;w,\lambda)$. Second, all of the primes dividing each denominator $b_{i,j}$ will be less than or equal to q_i, and q_i itself will divide $b_{i,j}$ at most to the $(k-j+1)$st power, so the denominators are becoming gradually smoother still.

Formally, we construct the fractions $a_{i,j}/b_{i,j}$ as follows. First, given $a_{i-1,k}/b_{i-1,k}$, where $1 \le i \le S$, we set $a_{i,1}/b_{i,1} = a_{i-1,k}/b_{i-1,k}$. Then, given $a_{i,j-1}/b_{i,j-1}$, where $2 \le j \le k$, we apply Lemma 3 with $p = q_i$, $l = k-j+1$, $N = q_i^{k-j+1}D(q_i)$, $c/d = a_{i,j-1}/b_{i,j-1}$, and $\mathcal{S} = \mathcal{A}(x,y;w,\lambda;q_i^{k-j+1})$. This gives us a subset of $\mathcal{A}(x,y;w,\lambda;q_i^{k-j+1})$, which we call $\mathcal{B}_{i,j}$, with cardinality less than q_i such that, if we define

$$\frac{a_{i,j}}{b_{i,j}} = \frac{a_{i,j-1}}{b_{i,j-1}} + \sum_{n\in\mathcal{B}_{i,j}} \frac{1}{n}$$

with $a_{i,j}/b_{i,j}$ in lowest terms, then $b_{i,j}$ divides $q_i^{k-j}D(q_i)$. Again, Lemma 8(c) justifies our use of Lemma 3 with these parameters. We also note that $D(q_{i-1}) = q_i^{k-1}D(q_i)$, and so our convention that $a_{i-1,k}/b_{i-1,k} = a_{i,1}/b_{i,1}$ is consistent with the divisibility property of the denominators.

We have gradually eliminated all prime factors not less than y' from the denominators of the fractions $a_{i,j}/b_{i,j}$; to finish the first stage of the proof, we eliminate powers of two as well. We set $a_{S+1,1}/b_{S+1,1} = a_{S,k}/b_{S,k}$ and define a sequence of fractions $a_{S+1,2}/b_{S+1,2}$, $a_{S+1,3}/b_{S+1,3}$, $\dots$, $a_{S+1,k}/b_{S+1,k}$, such that each $a_{S+1,j}/b_{S+1,j} - a_{S+1,j-1}/b_{S+1,j-1}$ is either zero or else is the reciprocal of an element of $\mathcal{A}(x,y';w,\lambda)$, and such that powers of two dividing the $b_{S+1,j}$ are diminishing. Given $a_{S+1,j-1}/b_{S+1,j-1}$, where $2 \le j \le k$, if 2^{k-j+1} exactly divides $b_{S+1,j-1}$, then we invoke Lemma 8(d) to choose an element $n_j \in \mathcal{A}(x,y';w,\lambda)$ that is exactly divisible by 2^{k-j+1} and set $a_{S+1,j}/b_{S+1,j} = a_{S+1,j-1}/b_{S+1,j-1} + 1/n_j$ and $\mathcal{B}_{S+1,j} = \{n_j\}$. Otherwise, we set $a_{S+1,j}/b_{S+1,j} = a_{S+1,j-1}/b_{S+1,j-1}$ and $\mathcal{B}_{S+1,j} = \emptyset$. It is easy to see inductively that, with this construction, $b_{S+1,j}$ divides $2^{k-j}D_0(y')$ for each $1 \le j \le k$.

Define

$$\mathcal{B} = \Bigl(\bigcup_{i=1}^{R} \mathcal{B}_i\Bigr) \cup \Bigl(\bigcup_{i=1}^{S+1}\bigcup_{j=2}^{k} \mathcal{B}_{i,j}\Bigr). \tag{15}$$

We note the largest prime factor of each element of a $\mathcal{B}_i$ is p_i, and the largest prime factor of each element of a $\mathcal{B}_{i,j}$ is q_i, except that the largest prime factor of each element of a $\mathcal{B}_{S+1,j}$ is less than y'. Furthermore, for a fixed $i \le S$, the various $\mathcal{B}_{i,j}$ have elements which are divisible by q_i to different powers; and the elements of the various $\mathcal{B}_{S+1,j}$ are divisible by different powers of two. We conclude that the union in the definition (15) of $\mathcal{B}$ is actually a disjoint union, and thus if we define $\mathcal{A} = \mathcal{A}(x,y;w,\lambda) \setminus \mathcal{B}$, then we have written

$$r = \frac{a_{S+1,k}}{b_{S+1,k}} + \sum_{n\in\mathcal{A}} \frac{1}{n} \tag{16}$$

with $b_{S+1,k}$ dividing $D_0(y')$. Moreover, the cardinality of $\mathcal{B}$ is bounded by

$$\sum_{w<p\le y}(p-1)+(k-1)\sum_{y'\le p\le w}(p-1)+(k-1)\ll y^2=x^{1-\varepsilon},$$

so that the cardinality of $\mathcal{A}$ is

$$\begin{aligned}|\mathcal{A}| &\ge |\mathcal{A}(x,y;w,\lambda)|-O(x^{1-\varepsilon})\\ &= (1-\lambda)\frac{x}{\zeta(k)}\rho\Big(\frac{2}{1-\varepsilon}\Big)\Big(1+O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big)\end{aligned}\tag{17}$$

by Lemma 8(a). We also note that

$$\frac{a_{S+1,k}}{b_{S+1,k}}=\frac{a_0}{b_0}+\sum_{n\in\mathcal{B}}\frac{1}{n}=\Big(\delta+O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big)+O\big(\,|\mathcal{B}|(\lambda x)^{-1}\big)=\delta+O\Big(\frac{\log_2^3 x}{\log y}\Big)\tag{18}$$

by equation (14) and the fact that $|\mathcal{B}|\ll x^{1-\varepsilon}$.

From the choice (13) of $\lambda=\lambda(\varepsilon,\delta,k)$ and the continuity of $\rho(u)$, we see that

$$\lim_{\substack{\varepsilon\to0\\ \delta\to0\\ k\to\infty}}\frac{1-\lambda}{\zeta(k)}\rho\Big(\frac{2}{1-\varepsilon}\Big)=\Big(1-\exp\Big(\frac{-r}{\rho(2)}\Big)\Big)\rho(2)=C(r).$$

We thus choose ε and δ sufficiently small and k sufficiently large, in terms of r and η, to ensure that $(1-\lambda)\rho(2/(1-\varepsilon))/\zeta(k)>C(r)-\eta$; then, from equation (17), we see that $|\mathcal{A}|>(C(r)-\eta)x$ for sufficiently large x. (We note that now x needs only to be sufficiently large in terms of r and η, since δ, ε, and k, and thus λ, have all been chosen in terms of r and η.) Moreover, all elements of $\mathcal{A}$ are certainly greater than λx. Therefore, to establish Theorem 1, it suffices by equation (16) to write $a_{S+1,k}/b_{S+1,k}$ as the sum of reciprocals of distinct integers not exceeding λx, without regard to the number of terms in the representation. This is the goal of the second stage of the proof.

We begin by applying to $a_{S+1,k}/b_{S+1,k}$ much the same process that we applied to r in the first stage of the proof. Recall that $y'=\exp(\log_2^4 x)$, and set $x'=(y')^{2k}$, $w'=y'$, and $y''=(\log x)/3k$. Let $q_1'>q_2'>\cdots>q_T'$ be the primes in $[y'',y')$, and for later consistency of notation, set $q_0'=q_S$. Since $\log x'/\log y'=2k$ and $\exp(\log_2^4 x')\le y''<y'=w'<(x')^{(1-\varepsilon)/k}$ for x sufficiently large, we may freely appeal to Lemmas 8 and 9 with the parameters x', y', and w' and any prime q_i'. In particular, by Lemma 9(c), we may choose λ' so that

$$0<\frac{a_{S+1,k}}{b_{S+1,k}}-\sum_{n\in\mathcal{A}_0(x',y';y',\lambda')}n^{-1}<(x')^{-1}\exp\Big(D(k)\frac{a_{S+1,k}}{b_{S+1,k}}\Big).\tag{19}$$

Define

$$\frac{a'_{0,k}}{b'_{0,k}}=\frac{a_{S+1,k}}{b_{S+1,k}}-\sum_{n\in\mathcal{A}_0(x',y';y',\lambda')}\frac{1}{n}$$

with $a'_{0,k}/b'_{0,k}$ in lowest terms, and notice that $b'_{0,k}$ divides $D_0(y')$. With this definition and the estimate (18) on the size of $a_{S+1,k}/b_{S+1,k}$, equation (19) becomes

$$0<\frac{a'_{0,k}}{b'_{0,k}}<(x')^{-1}\exp\bigg(D(k)\Big(\delta+O\Big(\frac{\log_2^3 x}{\log y}\Big)\Big)\bigg)\ll(x')^{-1}.$$

We recursively construct a sequence of fractions $a'_{1,1}/b'_{1,1}$, $a'_{1,2}/b'_{1,2}$, $\dots$, $a'_{1,k}/b'_{1,k}$, $a'_{2,1}/b'_{2,1}$, $\dots$, $a'_{T,k}/b'_{T,k}$ as follows. First, given $a'_{i-1,k}/b'_{i-1,k}$, where $1 \le i \le T$, we set $a'_{i,1}/b'_{i,1} = a'_{i-1,k}/b'_{i-1,k}$. Then, given $a'_{i,j-1}/b'_{i,j-1}$, where $2 \le j \le k$, we apply Lemma 3 with $p = q'_i$, $l = k-j+1$, $N = (q'_i)^{k-j+1}D(q'_i)$, $c/d = a'_{i,j-1}/b'_{i,j-1}$, and $\mathcal{S} = \mathcal{A}(x', y'; y', \lambda'; (q'_i)^{k-j+1})$. This gives us a subset of $\mathcal{A}(x', y'; y', \lambda'; (q'_i)^{k-j+1})$, which we call $\mathcal{B}'_{i,j}$, with cardinality less than q'_i such that, if we define

$$\frac{a'_{i,j}}{b'_{i,j}} = \frac{a'_{i,j-1}}{b'_{i,j-1}} + \sum_{n \in \mathcal{B}'_{i,j}} \frac{1}{n}$$

with $a'_{i,j}/b'_{i,j}$ in lowest terms, then $b'_{i,j}$ divides $(q'_i)^{k-j}D(q'_i)$.

Define

$$\mathcal{B}' = \bigcup_{i=1}^{T} \bigcup_{j=2}^{k} \mathcal{B}'_{i,j},$$

and notice that this is a disjoint union by the reasoning following equation (15) earlier. Thus, if we define $\mathcal{A}' = \mathcal{A}_0(x', y'; y', \lambda') \setminus \mathcal{B}'$, we see that

$$\frac{a_{S+1,k}}{b_{S+1,k}} = \frac{a'_{T,k}}{b'_{T,k}} + \sum_{n \in \mathcal{A}'} \frac{1}{n} \tag{20}$$

with $b'_{T,k}$ dividing $D_0(y'')$; in particular, $P(b'_{T,k}) < y''$. Moreover,

$$0 < \frac{a'_{T,k}}{b'_{T,k}} = \frac{a'_{0,k}}{b'_{0,k}} + \sum_{n \in \mathcal{A}'} \frac{1}{n} \ll (x')^{-1} + |\mathcal{A}'|(\lambda' x')^{-1} \ll (y')^2 (x')^{-1} = o((y'')^{-1}),$$

and so $a'_{T,k}/b'_{T,k} < (P(b'_{T,k}))^{-1}$ for x sufficiently large. Since $b'_{T,k}$ divides $D_0(y'')$, it is odd, and so we can apply Lemma 4 to find a set $\mathcal{C}$ of positive odd integers such that $a'_{T,k}/b'_{T,k} = \sum_{n \in \mathcal{C}} 1/n$ and

$$\max\{n \in \mathcal{C}\} \ll b'_{T,k} \prod_{p \le P(b'_{T,k})} p \le \prod_{p < y''} p^k = \exp\biggl(k \sum_{p < y''} \log p\biggr).$$

Chebyshev's bound for $\pi(t)$ implies that there is a real number $c < 2$ such that $\sum_{p<t} \log p < ct$ for any positive t. Therefore

$$\max\{n \in \mathcal{C}\} \ll \exp(kcy'') < x^{2/3}.$$

We have almost achieved our goal for the second stage of the proof, for

$$\frac{a_{S+1,k}}{b_{S+1,k}} = \sum_{n \in \mathcal{A}'} \frac{1}{n} + \sum_{n \in \mathcal{C}} \frac{1}{n};$$

but $\mathcal{A}'$ and $\mathcal{C}$ might not be disjoint. Define

$$\mathcal{D}_1 = \{n+1 : n \in \mathcal{A}' \cap \mathcal{C}\} \quad \text{and} \quad \mathcal{D}_2 = \{n(n+1) : n \in \mathcal{A}' \cap \mathcal{C}\}.$$

Since $1/n = 1/(n+1) + 1/(n(n+1))$, we have

$$\begin{aligned} \frac{a_{S+1,k}}{b_{S+1,k}} &= \sum_{n \in \mathcal{A}'} \frac{1}{n} + \sum_{n \in \mathcal{C} \setminus \mathcal{A}'} \frac{1}{n} + \sum_{n \in \mathcal{A}' \cap \mathcal{C}} \frac{1}{n} \\ &= \sum_{n \in \mathcal{A}'} \frac{1}{n} + \sum_{n \in \mathcal{C} \setminus \mathcal{A}'} \frac{1}{n} + \sum_{n \in \mathcal{D}_1} \frac{1}{n} + \sum_{n \in \mathcal{D}_2} \frac{1}{n}, \end{aligned} \tag{21}$$

and we claim that this is a disjoint representation of $a_{S+1,k}/b_{S+1,k}$ using denominators less than λx. We already know that $\max\{n \in \mathcal{A}'\} \le x'$ and $\max\{n \in \mathcal{C}\} \ll x^{2/3}$, and we easily see that

$$\max\{n \in \mathcal{D}_1 \cup \mathcal{D}_2\} \ll (\max\{n \in \mathcal{A}'\})^2 \le (x')^2,$$

so that the integers involved are of admissible size.

Clearly $\mathcal{A}'$ and $\mathcal{C} \setminus \mathcal{A}'$ are disjoint; and since the elements of $\mathcal{A}'$ and $\mathcal{C}$ are odd, those of $\mathcal{D}_1$ and $\mathcal{D}_2$ are even, and so each of the first two sets is disjoint from each of the last two. Finally, if there were an element n in $\mathcal{D}_1 \cap \mathcal{D}_2$, then there would exist m_1, $m_2 \in \mathcal{A}' \cap \mathcal{C}$ such that $n = m_1 + 1 = m_2(m_2 + 1)$. But then $m_1 \in \mathcal{A}' \subset \mathcal{A}_0(x', y'; y', \lambda')$ satisfies $m_1 = m_2^2 + m_2 - 1$, contradicting the definition of an $\mathcal{A}_0$-set. Therefore the four sets in the representation (21) are indeed disjoint, and Theorem 1 is established.

5. Prospects for improved results

Clearly the only barrier to establishing Theorem 1 in best-possible form is the presence of the factors $1 - \log 2$ in the expression for $C(r)$. This quantity arises as $\rho(2)$, which comes from using integers of size x that are roughly $x^{1/2}$-smooth, which in turn comes from the necessity that we have at least $p - 1$ multiples of every prime p at our disposal in order to invoke Lemma 2 and thus Lemma 3.

As noted in Section 2, the conclusion of Lemma 2 is best possible, since we can arrange for many sums of subsets to coincide. For randomly chosen sets, however, this behavior is very unlikely. Heuristically, a set of random nonzero elements of $\mathbf{Z}_p$ should have a subset summing to a randomly chosen target element of $\mathbf{Z}_p$ as soon as the size of the set is as large as a small power of $\log p$. We might expect, then, that if we were to attempt the construction in the proof of Theorem 1 using integers that were $x^{1-\varepsilon}$-smooth instead of $x^{(1-\varepsilon)/2}$-smooth, then we would have at least p^ε multiples of each prime p to choose from, and almost always we could find a small subset of those multiples to exclude to force the necessary cancellation of factors of p. If one could show that this were the case, the factors tending to $\rho(2)$ could be replaced by factors tending to $\rho(1) = 1$, and we would have established the best possible result.

Number theorists have also investigated whether positive rationals have Egyptian fraction representations of various specified forms. For instance, we have seen in Lemma 4 that any positive rational with odd denominator can be written as the sum of reciprocals of distinct odd positive integers (clearly no such representation exists for positive rationals whose denominator in reduced form is even, since the common denominator of such a representation will necessarily be odd). Graham [4] has proven a very general theorem showing that, for a certain class of subsets $\mathcal{Z}$ of the positive integers, any positive rational can be written as an Egyptian fraction with denominators restricted to elements of $\mathcal{Z}$, provided only that it satisfies two clearly necessary assumptions. Its denominator must not be divisible by any primes that don't divide any element of $\mathcal{Z}$, as in the case of odd denominators discussed above; and its size must be compatible with the sizes of the finite subsums of the reciprocals of $\mathcal{Z}$—the series of reciprocals of $\mathcal{Z}$ might converge, for instance, in which case large rationals could never be so represented.

It does not seem implausible, therefore, that any positive rational with odd denominator has an Egyptian fraction representation consisting of $\gg x$ unit fractions all of whose denominators are odd and at most x; or that a positive rational meeting

the local conditions prescribed by $\mathcal{Z}$ can be written as an Egyptian fraction using an asymptotically positive proportion of the elements of $\mathcal{Z}$. Furthermore, one might even believe that the best-possible bound for that proportion, which can be derived from the function $\sum_{n\in\mathcal{Z},\,\lambda x<n\le x} 1/n$ as we did in equation (1) for $\mathcal{Z}=\mathbf{Z}$, is in fact attainable.

References

[1] M. N. Bleicher, *A new algorithm for the expansion of Egyptian fractions*, J. Number Theory **4** (1972), 342–382. MR **48:**2052

[2] R. Breusch, *Solution to problem E4512*, American Math. Monthly **61** (1954), 200–201.

[3] S. W. Golomb, *An algebraic algorithm for the representation problems of the Ahmes papyrus*, Amer. Math. Monthly **69** (1962), 785–786.

[4] R. L. Graham, *On finite sums of unit fractions*, Proc. London Math. Soc. (3) **14** (1964), 193–207. MR **28:**3968

[5] A. Hildebrand, *On the number of positive integers $\le x$ and free of prime factors $> y$*, J. Number Theory **22** (1986), 289–307. MR **87d:**11066

[6] W. Sierpiński, *Sur les décompositions de nombres rationelles en fractions primaires*, Mathesis **65** (1956), 16–32. MR **17:**1185d

[7] R. C. Vaughan, *On a problem of Erdős, Straus and Schinzel*, Mathematika **17** (1970), 193–198. MR **44:**6600

[8] ———, *The Hardy–Littlewood method*, 2nd ed., Cambridge University Press, Cambridge, 1997. MR **98a:**11133

School of Mathematics, Institute for Advanced Study, Princeton, New Jersey 08540

Current address: Department of Mathematics, University of Toronto, Toronto, Canada M5S 3G3

E-mail address: gerg@math.toronto.edu