Journal of Number Theory **84**, 63–79 (2000)
doi:10.1006/jnth.2000.2508, available online at http://www.idealibrary.com on IDEAL®

Binary Egyptian Fractions

Ernest S. Croot III

Department of Mathematics, University of Georgia, Athens, Georgia 30609

David E. Dobbs

Department of Mathematics, University of Tennessee, Knoxville, Tennessee 37996-1300

John B. Friedlander

Department of Mathematics, University of Toronto, Ontario M5S 3G3, Canada

Andrew J. Hetzel

Department of Mathematics, University of Tennessee, Knoxville, Tennessee 37996-1300

and

Francesco Pappalardi

Dipartimento di Matematica, Università Roma Tre, Largo S. L. Murialdo 1, I-00146, Rome, Italy

Communicated by A. Granville

Received October 7, 1998; revised February 15, 1999

Let $A_k^*(n)$ be the number of positive integers a coprime to n such that the equation $a/n = 1/m_1 + \cdots + 1/m_k$ admits a solution in positive integers $(m_1, ..., m_k)$. We prove that the sum of $A_2^*(n)$ over $n \leqslant x$ is both $\gg x \log^3 x$ and also $\ll x \log^3 x$. For the corresponding sum where the a's are counted with multiplicity of the number of solutions we obtain the asymptotic formula. We also show that $A_k^*(n) \ll n^{\alpha_k + \varepsilon}$ where α_k is defined recursively by $\alpha_2 = 0$ and $\alpha_k = 1 - (1 - \alpha_{k-1})/(2 + \alpha_{k-1})$. © 2000 Academic Press

1. INTRODUCTION

An "Egyptian fraction representation" of a given rational a/n is a solution in positive integers of the equation

$$\frac{a}{n} = \frac{1}{m_1} + \cdots + \frac{1}{m_k}. \tag{1}$$

0022-314X/00 $35.00

In case $k=2$ we shall say it is a binary representation. A variety of questions about these representations have been posed and studied. Some of these require the m_i to be distinct but we shall not impose such a condition. We mention only a few of the many references and refer to the book by Guy [2] for a survey on this topic and a more extensive list of references.

As one example, a well-known conjecture of Erdős, Straus and Schinzel which is concerned with the "ternary" case states that when $k=3$ and $a=4$ the equation (1) has a solution for every given natural number $n>1$. It is easy to translate this into the problem of finding integer points on a family of rational surfaces. Evidence for this conjecture was provided by Vaughan [5], wherein, as an application of the large sieve it is shown that, if $E_a(x)$ counts the number of $n \leqslant x$ for which the equation $a/n = 1/m_1 + 1/m_2 + 1/m_3$ does not have a solution, then

$$E_a(x) \ll x \exp(-C(a) \log^{2/3} x).$$

Here $C(a)$ is a positive number depending at most on a. His result was later extended by Viola [6] to the case of general k.

In this note we change the point of view and instead of considering a fixed and n varying, we let n be fixed and vary a. To this purpose, let us set

$$A_k(n) = \#\{a, \text{ equation (1) has a solution}\}.$$

$$A_k^*(n) = \#\{a, (a,n)=1, \text{ equation (1) has a solution}\}.$$

In the binary case $k=2$ we shall suppress the subscript and simply write $A(n)$, $A^*(n)$. Let us note that trivially $A_k^*(n) \leqslant A_k(n) \leqslant kn$ and that

$$A_k(n) = \sum_{d \mid n} A_k^*(d). \tag{2}$$

Furthermore $A_k(n) \geqslant A_{k-1}(n)$, since any representation of length $k-1$ gives rise to one of length k by means of the trivial identity

$$\frac{1}{m} = \frac{1}{m+1} + \frac{1}{m(m+1)}. \tag{3}$$

For the most part we shall concentrate on the binary case. Even in this simplest case there does not seem to be very much known beyond two easily proven criteria which we record for completeness. The first of these, which is due to Bartoš [1], is not a very easily accessible reference.

Lemma 1. *Given $a, n \in \mathbf{N}$, the equation $a/n = 1/m_1 + 1/m_2$ has a solution for integers m_1, m_2 if and only if there exist positive integers k_1, k_2 such that*

$k_1k_2=n^2$, $a\,|\,n+k_1$, and $a\,|\,n+k_2$. In such a case, the solutions are $m_1=(n+k_1)/a$, $m_2=(n+k_2)/a$.

Proof. ($\Rightarrow$) Set $k_i=am_i-n$. It is clear that $a\,|\,n+k_i$ and also that

$$k_1k_2=(am_1-n)(am_2-n)=n^2+a(am_1m_2-n(m_1+m_2))=n^2.$$

($\Leftarrow$) It is enough to check that

$$\frac{a}{n+k_1}+\frac{a}{n+k_2}=a\frac{2n+k_1+k_2}{n^2+k_1k_2+n(k_1+k_2)}=\frac{a}{n}. \quad \blacksquare$$

We shall make repeated use of the second criterion, which has essentially been discovered by a number of authors, for instance Rav [3].

Lemma 2. *Consider fixed positive integers a, n such that $(a,n)=1$. Then there is a one-to-one correspondence between the solutions $(m_1,m_2)\in\mathbf{N}^2$ of the equation*

$$\frac{a}{n}=\frac{1}{m_1}+\frac{1}{m_2} \tag{4}$$

and the pairs $(u_1,u_2)\in\mathbf{N}^2$ with $(u_1,u_2)=1$, $u_1u_2\,|\,n$ and $a\,|\,u_1+u_2$. Furthermore if $(a,n)>1$ and such a pair (u_1,u_2) exists then the equation has a solution.

Proof of Lemma 2. Consider the map, say σ, which takes the solution (m_1,m_2) to the pair (u_1,u_2) defined by $u_i=m_i/(m_1,m_2)$ for $i=1,2$. Thus $(u_1,u_2)=1$ and it is clear that this pair cannot occur as the image of any other solution (for that fraction a/n). Moreover, we have

$$\frac{a}{n}=\frac{u_1+u_2}{(m_1,m_2)\,u_1u_2}.$$

Since $(a,n)=1$, we clearly have $a\,|\,u_1+u_2$. As $(u_1,u_2)=1$, it follows that $(u_1+u_2,u_1u_2)=1$, and so $u_1u_2\,|\,n$. Thus the image of σ lies inside the targeted set and it remains to show the map is surjective. Suppose then that there exist u_1,u_2 with the described properties. Write $n=cu_1u_2$ and $u_1+u_2=ab$ for suitable $b,c\in\mathbf{N}$. We have

$$\frac{a}{n}=\frac{u_1+u_2}{bcu_1u_2}=\frac{1}{bcu_1}+\frac{1}{bcu_2}.$$

Since the last argument applies irrespective of whether $(a,n)=1$ this completes the proof of the lemma. $\blacksquare$

From Lemma 2 we deduce:

COROLLARY 3. *For any* $\varepsilon > 0$ *we have*

$$A^*(n) \ll n^{\varepsilon} \qquad \textit{and} \qquad A(n) \ll n^{\varepsilon}. \tag{5}$$

Indeed, denoting as usual by $\tau(n)$ the number of positive integers dividing n, and using the well-known bound $\tau(n) \ll n^{\varepsilon}$, we have

$$A^*(n) \leqslant \sum_{\substack{u_1 \mid n,\, u_2 \mid n \\ (u_1, u_2) = 1}} \tau(u_1 + u_2) \ll n^{\varepsilon} \tau(n)^2 \ll n^{3\varepsilon}.$$

Furthermore, the second bound of (5) follows from the first since, by (2),

$$A(n) = \sum_{d \mid n} A^*(d) \ll n^{\varepsilon} \tau(n) \ll n^{2\varepsilon}.$$

Thus, the "probability" of a given proper fraction having a representation of length two is extremely small. In the final section we shall see that a weaker statement of the same nature holds in the case that "two" is replaced by any fixed k.

We remark that the above bounds can be sharpened slightly. Specifically, since $\log \tau(n) \ll \log n / \log \log n$, the above argument actually gives

$$\log A^*(n) \leqslant \log A(n) \ll \log n / \log \log n.$$

In this form the bound is best possible apart from the implied constant. Indeed, from the identity (3) it follows that every a dividing $n+1$ has a representation of type (4). Hence $A^*(n) \geqslant \tau(n+1)$ so $\log A^*(n) = \Omega(\log n / \log \log n)$.

On average the above estimates can be further improved. We shall show the following:

THEOREM 4. *We have the bounds*

$$x \log^3 x \ll \sum_{n \leqslant x} A^*(n) \ll x \log^3 x.$$

Using the theorem together with (2) and a little partial summation it is easy to give upper and lower bounds for the larger sum $\sum_{n \leqslant x} A(n)$ obtaining in each case the bound $x \log^4 x$.

It is (apparently) easier to evaluate the corresponding sum $\sum_{n \leqslant x} B^*(n)$ where $B^*(n) = \sum_a B^*(a, n)$ counts the number of solutions as a varies subject to $(a, n) = 1$ rather than just the number of such a for which solutions exist. This is because of Lemma 2 which reduces that problem to counting the pairs (u_1, u_2). If we normalize by counting only those solutions with $m_1 \leqslant m_2$ the result is

THEOREM 5.

$$\sum_{n \leqslant x} B^*(n) \sim \frac{C}{8} x \log^3 x,$$

where

$$C = \prod_{\substack{p \\ p \text{ prime}}} \left(1 - \frac{1}{p}\right)^2 \left(1 + \frac{2}{p}\right).$$

Throughout the paper p will always denote a prime. The proof we give yields the asymptotic formula with an error term which saves a factor $\log \log x$. By modifying the argument somewhat we could save a fixed power of $\log x$.

We remark that we were able to prove the upper and lower bounds of Theorem 4 with the explicit constants $C/8$ and $C/162$ respectively. It is obvious that Theorem 5 implies the upper bound in Theorem 4. Nevertheless, in the next section we shall give a direct proof by a much easier argument.

Much of the early work on this paper was accomplished while D.D. and J.F. were visiting the Department of Mathematics, Università Roma Tre with partial support of G.N.S.A.G.A., Consiglio Nazionale delle Ricerche. D.D. and A.H. were also supported in part by NSF/REU, and J.F. by NSERC Canada.

2. PROOF OF THE UPPER BOUND

In this section we prove that

$$\sum_{n \leqslant x} A^*(n) \ll x \log^3 x.$$

Interchanging the order of summation, we have by Lemma 2

$$\sum_{n \leqslant x} A^*(n) \leqslant \sum_{a \leqslant 2x} \sum_{\substack{u_1 u_2 \leqslant x \\ (u_1, u_2) = 1, \\ u_1 + u_2 \equiv 0 \ (\mathrm{mod}\ a)}} \sum_{\substack{n \leqslant x \\ (n, a) = 1, \\ n \equiv 0 \ (\mathrm{mod}\ u_1 u_2)}} 1.$$

Even after the condition $(n, a) = 1$ is discarded the innermost sum is $\leqslant x/u_1 u_2$. Hence, discarding also the condition $(u_1, u_2) = 1$, we have

$$\sum_{n \leqslant x} A^*(n) \leqslant 2x \sum_{a \leqslant 2x} \sum_{u_1 \leqslant x} \frac{1}{u_1} \sum_{\substack{u_1 \leqslant u_2 \leqslant x/u_1 \\ u_2 \equiv -u_1 \pmod{a}}} \frac{1}{u_2},$$

where we have labeled so that $u_1 \leqslant u_2$ (doubling to compensate). Therefore, since $u_1 + u_2 \equiv 0 \pmod{a}$ and $u_2 \geqslant u_1$, we have $u_2 \geqslant a/2$.

Thus, the inner sum over u_2 is

$$\leqslant \sum_{0 \leqslant j \leqslant x} \frac{1}{(1/2)\, a + ja} \ll \frac{1}{a} \log x$$

and so

$$\sum_{n \leqslant x} A^*(n) \ll x \sum_{a \leqslant 2x} \frac{1}{a} \sum_{u_1 \leqslant x} \frac{1}{u_1} \log x \ll x \log^3 x. \quad \blacksquare$$

3. PROOF OF THE LOWER BOUND

It is also very easy to give a lower bound but, unlike the previous argument, this one does not lead to the right order of magnitude. Specifically we have

Proposition 6. $\sum_{n \leqslant x} A^*(n) \gg x \log x$.

Proof. Take $u_1 = p$ prime, say with $x^{1/3} \leqslant p \leqslant x^{1/2}$. Consider those pairs (u_1, u_2) with $u_1 u_2 = n$ and any $a \mid u_1 + u_2$ (other than $a = 1$). We have

$$\sum_{n \leqslant x} A^*(n) \geqslant \sum_{x^{1/3} \leqslant p \leqslant x^{1/2}} \sum_{\substack{m \leqslant x/p \\ m \not\equiv 0 \pmod{p}}} (\tau(p+m) - 1)$$

$$\gg x \log x \sum_{x^{1/3} \leqslant p \leqslant x^{1/2}} \frac{1}{p} \gg x \log x. \quad \blacksquare$$

Alternatively this could be deduced by summing over n the lower bound $A^*(n) \geqslant \tau(n+1)$. We were able to sharpen the bound to

$$\sum_{n \leqslant x} A^*(n) \gg x \frac{\log^2 x}{\log \log x},$$

using an elaboration of the above simple idea. However the proof is not so brief, yet the result still falls short of the right order. Hence we do not give it here.

The problem with using Lemma 2 to find a lower bound is that some of the pairs (a, n) give rise to more than one pair u_1, u_2. One might expect that this multiplicity is, for $a > n^\varepsilon$, almost always 0 or 1 and therefore that

$$\sum_{n \leqslant x} A^*(n) \sim \sum_{n \leqslant x} B^*(n).$$

This would solve our problem in view of our result for the latter sum.

Perhaps this can be proven. Certainly if a is sufficiently large there is no such multiplicity and this fact will lead us to our lower bound in Theorem 4. Specifically, we have:

LEMMA 7. *Suppose* $(a, n) = 1$ *and* $a > 2n^{2/3}$. *Then, there can be at most one solution to*

$$\frac{a}{n} = \frac{1}{m_1} + \frac{1}{m_2}, \tag{6}$$

where $m_1 \leqslant m_2$ *are positive integers.*

Proof. First observe that any solution to (6) with $m_1 \leqslant m_2$ must have $m_1 < n^{1/3}$ since

$$\frac{2}{n^{1/3}} < \frac{a}{n} \leqslant \frac{2}{m_1}.$$

But then we must have that $m_2 > n^{2/3}$, since n divides $m_1 m_2$ and hence is no larger than it.

Now suppose there is another solution (m_1', m_2') to (6). Then, $m_1' \neq m_1$, $m_1' < n^{1/3}$ and $m_2' > n^{2/3}$, and so

$$\frac{1}{n^{2/3}} < \frac{1}{m_1 m_1'} \leqslant \left| \frac{1}{m_1} - \frac{1}{m_1'} \right| = \left| \frac{1}{m_2} - \frac{1}{m_2'} \right| < \frac{1}{n^{2/3}}$$

which is impossible. Therefore there can be no second solution (m_1', m_2') to (6). ∎

We shall also need several special cases of the following result.

LEMMA 8. *Suppose that* f *is a non-negative multiplicative function which takes the form*

$$f(n) = \prod_{p \mid n} \left(1 + \frac{c(p)}{p} + \frac{c(p^2)}{p^2} + \cdots \right),$$

where

$$|c(p^r)| \leqslant B^r, \tag{7}$$

for all but finitely many primes p. Then,

$$\sum_{\substack{j \leqslant y \\ (j,b)=1}} f(ja) = f(a)\,\frac{\varphi(b)}{b} \prod_{p \nmid ab} \left(1 + \frac{f(p)-1}{p}\right) y + O(f(a)\,\tau(b)\log^B y),$$

where the implied constant depends at most on f and B. We also have, for any integer $r \geqslant 0$,

$$\sum_{\substack{j \leqslant y \\ (j,b)=1}} \frac{f(ja)}{ja} (\log ja)^r = \frac{f(a)}{a}\,\frac{\varphi(b)}{b} \prod_{p \nmid ab} \left(1 + \frac{f(p)-1}{p}\right) \int_a^{ay} (\log v)^r \frac{dv}{v}$$
$$+ O\left(\frac{f(a)}{a} ((\log ay)^r + (\log(\tau(b)\log y))^{r+1})\right),$$

where now the implied constant may also depend on r.

Proof. It is possible to prove a result of this type using contour integration and this would enable us to sharpen some of our error terms. However we give an elementary proof more in keeping with the spirit of this work.

Define the multiplicative function $g(n) = \prod_{p \mid n} (f(p) - 1)$ so that

$$f(n) = \sum_{s \mid n} \mu^2(s)\, g(s)$$

and we have

$$\begin{aligned}
\sum_{\substack{j \leqslant y \\ (j,b)=1}} f(ja) &= \sum_{\substack{j \leqslant y \\ (j,b)=1}} \sum_{s \mid ja} \mu^2(s)\, g(s) \\
&= \sum_{\substack{j \leqslant y \\ (j,b)=1}} \sum_{s \mid a} \mu^2(s)\, g(s) \sum_{\substack{t \mid j \\ (t,a)=1}} \mu^2(t)\, g(t) \\
&= f(a) \sum_{\substack{j \leqslant y \\ (j,b)=1}} \sum_{\substack{t \mid j \\ (t,a)=1}} \mu^2(t)\, g(t) \\
&= f(a) \sum_{\substack{t \leqslant y \\ (t,ab)=1}} \mu^2(t)\, g(t) \sum_{\substack{j \leqslant y \\ t \mid j \\ (j,b)=1}} 1.
\end{aligned}$$

To the innermost sum we apply the well-known elementary formula

$$\sum_{\substack{n \leqslant y \\ (n, b) = 1}} 1 = y \frac{\varphi(b)}{b} + O(\min(\tau(b), y)), \tag{8}$$

obtaining

$$\sum_{\substack{j \leqslant y \\ (j, b) = 1}} f(ja) = f(a) \frac{\varphi(b)}{b} y \sum_{\substack{t \leqslant y \\ (t, ab) = 1}} \mu^2(t) \frac{g(t)}{t}$$

$$+ O\Bigg(f(a)\, \tau(b) \sum_{\substack{t \leqslant y \\ (t, ab) = 1}} \mu^2(t)\, g(t) \Bigg). \tag{9}$$

Let P denote the set of those exceptional primes not satisfying (7), together with those primes for which $p \leqslant B$. Then P is a finite set and

$$\sum_{p \in P} \log(1 + |g(p)|) \ll_{f, B} 1.$$

Thus

$$\log \Bigg| \sum_{t \leqslant x} \mu^2(t)\, g(t) \Bigg| \leqslant \log \prod_{p \leqslant x} \left(1 + \frac{B}{p} + \frac{B^2}{p^2} + \cdots \right) + \sum_{p \in P} \log(1 + |g(p)|)$$

$$= \sum_{p \leqslant x} \frac{B}{p} + O(1) = B \log \log x + O(1)$$

so we find, on exponentiating both sides, that

$$\sum_{t \leqslant x} \mu^2(t)\, g(t) \ll \log^B x. \tag{10}$$

Also, for t a positive squarefree integer,

$$|g(t)| \leqslant \prod_{p \in P} |g(p)| \prod_{p \mid t,\, p \notin P} \left(\frac{B}{p} + \frac{B^2}{p^2} + \cdots \right)$$

$$\ll B^{\nu(t)} \prod_{p \mid t,\, p \notin P} (p - B)^{-1} \ll_{B, \varepsilon} \frac{1}{t^{1-\varepsilon}},$$

where $\nu(t) = \sum_{p \mid t} 1$. Thus,

$$\sum_{\substack{t = 1 \\ (t, ab) = 1}}^{\infty} \mu^2(t) \frac{g(t)}{t} \text{ converges,}$$

and so

$$\sum_{\substack{t \leqslant y \\ (t, ab)=1}} \mu^2(t) \frac{g(t)}{t} = \sum_{\substack{t=1 \\ (t, ab)=1}}^{\infty} \mu^2(t) \frac{g(t)}{t} - \sum_{\substack{t > y \\ (t, ab)=1}} \mu^2(t) \frac{g(t)}{t}$$

$$= \prod_{p \nmid ab} \left(1 + \frac{g(p)}{p}\right) + O\left(\sum_{t > y} \mu^2(t) \frac{g(t)}{t}\right). \tag{11}$$

From (10) and partial summation we have that

$$\sum_{t > y} \mu^2(t) \frac{g(t)}{t} \ll \frac{\log^B y}{y}.$$

Combining this with (9) and (11), we find that

$$\sum_{\substack{j \leqslant y \\ (j, b)=1}} f(ja) = f(a) \frac{\varphi(b)}{b} \prod_{p \nmid ab} \left(1 + \frac{g(p)}{p}\right) y + O(f(a)\, \tau(b) \log^B y)$$

giving the first result of the lemma.

The second statement follows from the first by partial summation. For the smallest part of the range, say $t \leqslant T$, we first use the trivial upper bound provided by

$$\sum_{\substack{j \leqslant t \\ (j, b)=1}} f(ja) \leqslant \sum_{j \leqslant t} f(ja)$$

and then apply the first statement of the lemma and partial summation to the latter sum. For the bulk of the range, $T \leqslant t \leqslant y$, we apply partial summation in the usual fashion. Choosing $T = \tau(b)$ which is close to optimal we obtain the result. ∎

We are now ready to prove the lower bound in Theorem 4. From Lemma 7 we can see immediately that for any $\delta > 0$

$$\sum_{n \leqslant x} A^*(n) \geqslant \sum_{2x^{2/3} < a < x^{1-3\delta}} \sum_{\substack{u_1 u_2 < x^{1-\delta},\, u_1 \leqslant u_2 \\ (u_1, u_2)=1 \\ u_1 + u_2 \equiv 0 \pmod a}} \sum_{\substack{n \leqslant x \\ (n, a)=1 \\ n \equiv 0 \pmod{u_1 u_2}}} 1.$$

Actually the lemma implies the stronger statement with $\delta = 0$ but it will be convenient in what follows to take δ small and positive thus restricting ourselves to a smaller range of the variables.

We estimate the innermost sum using Lemma 8. We find that

$$\sum_{n\leqslant x} A^*(n) \gg x \sum_{2x^{2/3}<a<x^{1-3\delta}} \frac{\varphi(a)}{a} \sum_{\substack{u_1\leqslant x^{1-2\delta}\\(u_1,a)=1}} \frac{1}{u_1} \sum_{\substack{u_2\leqslant x^{1-\delta}/u_1\\(u_2,u_1)=1\\u_2\equiv -u_1 \ (\mathrm{mod}\ a)}} \frac{1}{u_2}$$

and, writing $u_1+u_2=ma$ we have

$$\sum_{n\leqslant x} A^*(n) \gg x \sum_{2x^{2/3}<a<x^{1-3\delta}} \frac{\varphi(a)}{a^2} \sum_{\substack{u_1\leqslant x^{\delta}\\(u_1,a)=1}} \frac{1}{u_1} \sum_{\substack{1\leqslant m\leqslant x^{\delta}\\(m,u_1)=1}} \frac{1}{m}. \tag{12}$$

Using again Lemma 8 we may bound below the innermost sum in (12) obtaining

$$\sum_{\substack{1\leqslant m\leqslant x^{\delta}\\(m,u_1)=1}} \frac{1}{m} \gg_{\delta} \frac{\varphi(u_1)}{u_1}\log x$$

and hence

$$\sum_{n\leqslant x} A^*(n) \gg_{\delta} x\log x \sum_{2x^{2/3}<a<x^{1-3\delta}} \frac{\varphi(a)}{a^2} \sum_{\substack{u_1\leqslant x^{\delta}\\(u_1,a)=1}} \frac{\varphi(u_1)}{u_1^2}.$$

For this sum we may again use Lemma 8 to obtain

$$\sum_{\substack{u_1\leqslant x^{\delta}\\(u_1,a)=1}} \frac{\varphi(u_1)}{u_1^2} \gg \frac{\varphi(a)}{a}\log y,$$

and thus,

$$\sum_{n\leqslant x} A^*(n) \gg_{\delta} x\log^2 x \sum_{2x^{2/3}<a<x^{1-3\delta}} \frac{\varphi^2(a)}{a^3}.$$

To this last sum we again apply Lemma 8 to deduce that

$$\sum_{2x^{2/3}<a<x^{1-3\delta}} \frac{\varphi^2(a)}{a^3} \gg_{\delta} \log x,$$

and so, as claimed, that

$$\sum_{n\leqslant x} A^*(n) \gg x\log^3 x.$$

4. COUNTING WITH MULTIPLICITY

In this section we give the proof of Theorem 5. More precisely we prove the following general result from which Theorem 5 follows as the special case $A = 2x$.

PROPOSITION 9. *If* $A \leqslant \sqrt{2x}$, *then*

$$\sum_{n \leqslant x} \sum_{\substack{a \leqslant A \\ (a,n)=1}} B^*(a,n) = \frac{C}{2} x \left(\frac{1}{2} \log^2 x \log A - \frac{1}{3} \log^3 A \right) + O\left(x \frac{\log^3 x}{\log \log x} \right), \tag{13}$$

and if $\sqrt{2x} \leqslant A \leqslant 2x$, *then*

$$\sum_{n \leqslant x} \sum_{\substack{a \leqslant A \\ (a,n)=1}} B^*(a,n) = \frac{C}{2} x \left(-\frac{\log^3 x}{12} + \log x \log A \log(x/A) + \frac{\log^3 A}{3} \right) + O\left(x \frac{\log^3 x}{\log \log x} \right). \tag{14}$$

Here C is, as stated earlier, given by

$$C = \prod_p \left(1 - \frac{1}{p}\right)^2 \left(1 + \frac{2}{p}\right),$$

and the result can be refined to give an error term $O(x \log^M x)$ for some $M < 3$, perhaps $M = 2$.

Proof. First suppose $A \leqslant \sqrt{2x}$. To prove (13) we shall estimate

$$\sum_{n \leqslant x} \sum_{\substack{a \leqslant A \\ (a,n)=1}} B^*(a,n) = \sum_{a \leqslant A} \sum_{\substack{u_1 u_2 \leqslant x,\, u_1 \leqslant u_2 \\ (u_1, u_2)=1 \\ u_1 + u_2 \equiv 0 \pmod a}} \sum_{\substack{n \leqslant x \\ u_1 u_2 \mid n \\ (n,a)=1}} 1.$$

The inner sum is equal to

$$\frac{x}{u_1 u_2} \frac{\varphi(a)}{a} + O(\min(\tau(a), x/(u_1 u_2)).$$

A little calculation shows that this error term, when summed over a, u_1, and u_2, contributes $O(x \log^2 x)$. Thus,

$$\sum_{n \leqslant x} \sum_{\substack{a \leqslant A \\ (a,n)=1}} B^*(a,n) = x \sum_{a \leqslant A} \frac{\varphi(a)}{a} \sum_{\substack{u_1 u_2 \leqslant x,\, u_1 \leqslant u_2 \\ (u_1,u_2)=1 \\ u_1+u_2 \equiv 0 \pmod{a}}} \frac{1}{u_1 u_2} + O(x \log^2 x).$$

We make the change of variable from u_2 to j where $u_2 = ja - u_1$ and then split the range of j, obtaining for the inner sum over u_1, u_2

$$\sum_{\sqrt{2x}/a < j \leqslant x/a} \sum_{\substack{u_1(ja-u_1) \leqslant x \\ (u_1, ja)=1}} \frac{1}{u_1(ja-u_1)} + \sum_{1 \leqslant j \leqslant \sqrt{2x}/a} \sum_{\substack{u_1 \leqslant ja/2 \\ (u_1, ja)=1}} \frac{1}{u_1(ja-u_1)} + O\left(\frac{1}{a}\right).$$

In these sums we replace $ja - u_1$ by ja, estimate the error so obtained when summed over the variables; it is $O(x \log^2 x)$. This gives

$$\sum_{n \leqslant x} \sum_{\substack{a \leqslant A \\ (a,n)=1}} B^*(a,n) = xS_1 + xS_2 + O(x \log^2 x), \tag{15}$$

where

$$S_1 = \sum_{a \leqslant A} \frac{\varphi(a)}{a} \sum_{\sqrt{2x}/a < j \leqslant x/a} \frac{1}{ja} \sum_{\substack{u_1 \leqslant x/(ja) \\ (u_1, ja)=1}} \frac{1}{u_1} \tag{16}$$

and

$$S_2 = \sum_{a \leqslant A} \frac{\varphi(a)}{a} \sum_{1 \leqslant j \leqslant \sqrt{2x}/a} \frac{1}{ja} \sum_{\substack{u_1 \leqslant ja/2 \\ (u_1, ja)=1}} \frac{1}{u_1}. \tag{17}$$

From Eq. (8), supplemented by a trivial bound, one deduces using partial summation that

$$\sum_{\substack{u_1 \leqslant y \\ (u_1, ja)=1}} \frac{1}{u_1} = \frac{\varphi(ja)}{ja} \log y + O(1 + \log \tau(ja)). \tag{18}$$

We use this to estimate the inner sum in both S_1 and S_2. In both instances the contribution of the error terms, when summed over j and a, is bounded by

$$\ll x \sum_{a \leqslant A} \frac{\varphi(a)}{a} \sum_{j \leqslant x/a} \frac{\log(\tau(ja))}{ja} \ll x \frac{\log^3 x}{\log \log x}.$$

We remark that in this last step we have used the trivial bound for the divisor function $\log(\tau(m)) \ll \log m/\log\log m$ whereas we really need only a bound for this on average. It is this step which in the current argument sets the limit for the error term in the proposition (and the theorem).

We are now left to estimate

$$S_1' + S_2' = \sum_{a \leqslant A} \frac{\varphi(a)}{a} \sum_{\sqrt{2x/a} < j \leqslant x/a} \frac{1}{ja} \frac{\varphi(ja)}{ja} \log(x/ja) + \sum_{a \leqslant A} \frac{\varphi(a)}{a} \sum_{1 \leqslant j \leqslant \sqrt{2x/a}} \frac{1}{ja} \frac{\varphi(ja)}{ja} \log(ja/2).$$

By Lemma 8 we obtain that

$$\sum_{\sqrt{2x/a} < j \leqslant x/a} \frac{1}{ja} \frac{\varphi(ja)}{ja} \log(x/ja) = \frac{c_a}{8a} \log^2 x + O\left(\frac{\log^2 x}{a \log\log x}\right), \tag{19}$$

where

$$c_a = \frac{6}{\pi^2} \prod_{p \mid a} \left(1 + \frac{1}{p}\right)^{-1},$$

and also that

$$\sum_{1 \leqslant j \leqslant \sqrt{2x/a}} \frac{1}{ja} \frac{\varphi(ja)}{ja} \log(ja/2) = \frac{c_a}{2a} \left(\frac{\log^2 x}{4} - \log^2 a\right) + O\left(\frac{\log^2 x}{a \log\log x}\right).$$

Thus, making these substitutions above, we have

$$\sum_{n \leqslant x} \sum_{\substack{a \leqslant A \\ (a,n)=1}} B^*(a,n) = x \sum_{a \leqslant A} c_a \frac{\varphi(a)}{a^2} \left(\frac{\log^2 x}{4} - \frac{\log^2 a}{2}\right) + O\left(x \frac{\log^3 x}{\log\log x}\right). \tag{20}$$

Applying again Lemma 8 we have

$$\sum_{a \leqslant A} c_a \frac{\varphi(a)}{a^2} \log^r a = C \frac{\log^{r+1} A}{r+1} + O\left(\frac{\log^{r+1} x}{\log\log x}\right), \tag{21}$$

where C is as before. Applying this for the cases $r = 0$ and $r = 2$ we deduce from (20) that

$$\sum_{n \leqslant x} \sum_{\substack{a \leqslant A \\ (a,n)=1}} B^*(a,n) = \frac{C}{2} x \left(\frac{1}{2} \log^2 x \log A - \frac{1}{3} \log^3 A\right) + O\left(x \frac{\log^3 x}{\log\log x}\right).$$

Now suppose that $\sqrt{2x} < A \leqslant 2x$. To prove the propositon in this case we must estimate

$$\sum_{n \leqslant x} \sum_{\substack{a \leqslant A \\ (a,n)=1}} B^*(a,n) = \sum_{n \leqslant x} \sum_{\substack{1 \leqslant a \leqslant \sqrt{2x} \\ (a,n)=1}} B^*(a,n) + \sum_{n \leqslant x} \sum_{\substack{\sqrt{2x} < a \leqslant A \\ (a,n)=1}} B^*(a,n).$$

For the former of these we have

$$\sum_{n \leqslant x} \sum_{\substack{1 \leqslant a \leqslant \sqrt{2x} \\ (a,n)=1}} B^*(a,n) = \frac{5C}{48} x \log^3 x + O\left(x \frac{\log^3 x}{\log\log x}\right) \tag{22}$$

by the result for the earlier case and it remains to treat the sum

$$\sum_{n \leqslant x} \sum_{\substack{\sqrt{2x} < a \leqslant A \\ (a,n)=1}} B^*(a,n) = \sum_{\sqrt{2x} < a \leqslant A} \sum_{\substack{u_1 u_2 \leqslant x,\, u_1 \leqslant u_2 \\ (u_1, u_2)=1 \\ u_1 + u_2 \equiv 0 \pmod a}} \sum_{\substack{n \leqslant x \\ u_1 u_2 \mid n \\ (n,a)=1}} 1.$$

We can estimate this in a way analogous to the previous case, and here it is somewhat simpler since the sum

$$\sum_{1 \leqslant j \leqslant \sqrt{2x}/a} \sum_{\substack{u_1 \leqslant ja/2 \\ (u_1, ja)=1}} \frac{1}{u_1(ja - u_1)},$$

which appeared earlier, vanishes when $a > \sqrt{2x}$. One can show, in exactly the same way as before, that

$$\sum_{n \leqslant x} \sum_{\substack{\sqrt{2x} < a < A \\ (a,n)=1}} B^*(a,n)$$
$$= x \sum_{\sqrt{2x} < a \leqslant A} \frac{\varphi(a)}{a} \sum_{1 \leqslant j \leqslant x/a} \frac{1}{ja} \sum_{\substack{u_1 \leqslant x/(ja) \\ (u_1, ja)=1}} \frac{1}{u_1} + O(x \log^2 x).$$

Here the inner sum was estimated in Eq. (18). From this we obtain that the right side above is

$$= x \sum_{\sqrt{2x} < a \leqslant A} \frac{\varphi(a)}{a} \sum_{1 \leqslant j \leqslant x/a} \frac{\varphi(ja)}{(ja)^2} \log(x/(ja)) + O\left(x \frac{\log^3 x}{\log\log x}\right).$$

The inner sum here is similar to (19) and by Lemma 8 is equal to

$$\frac{c_a}{2a} \log^2(x/a) + O\left(\frac{\log^2 x}{a \log\log x}\right),$$

where c_a is as defined earlier. Thus we have

$$\sum_{n\leqslant x}\sum_{\substack{\sqrt{2x}<a\leqslant A\\(a,n)=1}} B^*(a,n)=x\sum_{\sqrt{2x}<a\leqslant A}\frac{c_a}{2}\frac{\varphi(a)}{a^2}\log^2(x/a)+O\left(x\frac{\log^3 x}{\log\log x}\right),$$

which by (21) is equal to

$$\frac{C}{2}x\left(-\frac{7}{24}\log^3 x+\frac{\log^3 A}{3}+\log x\log A\log(x/A)\right)+O\left(x\frac{\log^3 x}{\log\log x}\right).$$

By combining this with (22) we complete the proof of the proposition.

5. THE CASE $k>2$

Lacking in this case a characterization comparable to that given in Lemma 2 we were not even able to prove that $A_k^*(n)\ll n^{\varepsilon}$ for $k>2$ although it is reasonable to suspect that this might be the right estimate. We can prove the following much weaker result:

PROPOSITION 10. *Let $k>1$ be an integer and let α_k be the sequence defined recursively by $\alpha_2=0$, $\alpha_k=1-(1-\alpha_{k-1})/(2+\alpha_{k-1})$ so that $\alpha_3=1/2$, $\alpha_4=4/5$, $\alpha_5=13/14$, ... Then for every $\varepsilon>0$,*

$$A_k(n)\ll n^{\alpha_k+\varepsilon}$$

where the implied constant depends on k and ε.

Proof. We have already proven the result for $k=2$ in (5), so we next assume that the conclusion holds for $k-1$. If $a\in\mathbf{N}$ is such that $a/n=1/m_1+\cdots+1/m_k$ where, without loss of generality, $m_1\leqslant m_2\leqslant\cdots\leqslant m_k$ then it is easy to verify that $n/a<m_1\leqslant kn/a$.

Let us fix some $\beta>0$ to be determined and consider an integer $a>n^{\beta}$. By the preceding argument, we have that $m_1\leqslant kn^{1-\beta}$. Therefore, applying the inductive hypothesis to $(m_1a-n)/m_1n=1/m_2+\cdots+1/m_k$, we obtain

$$\begin{aligned}A_k(n)&\leqslant\sum_{m_1\leqslant kn^{1-\beta}}A_{k-1}(nm_1)+n^{\beta}\\&\ll_{k,\varepsilon}n^{\alpha_{k-1}+\varepsilon/2}\sum_{m_1\leqslant kn^{1-\beta}}m_1^{\alpha_{k-1}+\varepsilon/2}+n^{\beta}\\&\ll_{k,\varepsilon}n^{\alpha_{k-1}+\varepsilon/2}\cdot n^{(1-\beta)(1+\alpha_{k-1}+\varepsilon/2)}+n^{\beta}.\end{aligned}$$

We can arrange that the right-hand side is $\ll_{k,\varepsilon} n^{\alpha_k+\varepsilon}$ by choosing

$$\alpha_k = \beta = \alpha_{k-1} + (1-\beta)(1+\alpha_{k-1}),$$

that is $\alpha_k = (1+2\alpha_{k-1})/(2+\alpha_{k-1})$. The assertion follows. ∎

We remark that it would not be difficult to adapt the proof of the upper bound to show that

$$\sum_{n \leqslant x} A_k^*(n) \ll x^{1+\alpha_k} \log^{\beta_k} x$$

with an appropriate value β_k. However this is still rather far from the expected order of magnitude.

REFERENCES

1. P. Bartoš, A remark on the number of solutions of the equation $1/x + 1/y = a/b$ in natural numbers, *Časopis Pěst. Mat.* **95** (1970), 411–415.
2. R. K. Guy, "Unsolved Problems in Number Theory," 2nd ed., Springer-Verlag, New York, 1994.
3. Y. Rav, On the representation of rational numbers as a sum of a fixed number of unit fractions, *J. Reine Angew. Math.* **222** (1966), 207–213.
4. E. C. Titchmarsh, "The Riemann Zeta-Function," 2nd ed., revised by D. R. Heath-Brown, Oxford, 1986.
5. R. C. Vaughan, On a problem of Erdős, Straus and Schinzel, *Mathematika* **17** (1970), 193–198.
6. C. Viola, On the diophantine equations $\Pi_0^k x_i - \sum_0^k x_i = n$ and $\sum_0^k 1/x_i = a/n$, *Acta Arith.* **22** (1972/73), 339–352.