Denser Egyptian fractions

by

GREG MARTIN (Toronto)

1. Introduction. An *Egyptian fraction* is a sum of reciprocals of distinct positive integers, so called because the ancient Egyptians represented rational numbers in that way. Among the many questions about Egyptian fraction representations of rational numbers that have been posed by Erdős and Graham (see [4] as well as [5, Section D11]) are some concerning representations where the sizes of the denominators are bounded above relative to the number of terms, and these questions have been the focus of several papers in the last decade. For instance, the author showed in [8] that every positive rational number r has Egyptian fraction representations where the number of terms is of the same order of magnitude as the largest denominator, improving a result of Yokota [10]. More precisely, for each positive integer t let us define

$$\mathcal{H}_t(r) = \left\{ (x_1, \ldots, x_t) \in \mathbb{Z}^t : x_1 > \ldots > x_t \geq 1,\ \sum_{i=1}^{t} \frac{1}{x_i} = r \right\},$$

the collection of all sets of denominators in t-term Egyptian fraction representations of r, and

$$(1) \qquad M_t(r) = \inf\{x_1 : (x_1, \ldots, x_t) \in \mathcal{H}_t(r)\},$$

the smallest integer x that is the largest denominator in a t-term Egyptian fraction representation of r (unless no such t-term representation exists in which case $M_t(r)$ equals infinity). In [8] it was established that for every positive rational number r, there is an infinite sequence of integers t such that $M_t(r) \leq C(r)t$, where $C(r)$ is a certain constant.

In this paper we improve this result, and indeed we completely determine the asymptotic size of $M_r(t)$ for all positive rational numbers r and integers t. Because the set $\mathcal{H}_t(r)$ might well be empty for small values of t, we introduce the notation $t_0(r)$ for the least number of terms in any Egyptian

---

fraction representation of r. Note that $\mathcal{H}_t(r)$ is nonempty for every $t \geq t_0(r)$, since a representation with t terms can be converted into one with $t+1$ terms by "splitting" the term with largest denominator, using the identity

$$(2) \qquad \frac{1}{n} = \frac{1}{n+1} + \frac{1}{n(n+1)}.$$

Therefore $M_t(r) = \infty$ for $t < t_0(r)$ and $M_t(r) < \infty$ for $t \geq t_0(r)$. (The only flaw in this argument arises when the largest denominator is $n = 1$, in which case the splitting identity does not yield distinct terms; and in fact, the rational number 1 has an Egyptian fraction representation with one term but no representation with two terms. For this reason, we make the convention that $t_0(1) = 3$.) We can now state the following theorem:

THEOREM 1. *For all positive rational numbers r and all integers $t \geq t_0(r)$, we have*

$$(3) \qquad M_t(r) = \frac{t}{1 - e^{-r}} + O_r\left(\frac{t \log \log 3t}{\log 3t}\right).$$

Furthermore, this is best possible, in that the order of magnitude of the error term cannot be reduced.

The methods used to establish Theorem 1 allow us also to address another problem posed in [4]. Erdős and Graham observe that a prime power can never be the largest denominator in an Egyptian fraction representation of 1, nor can a tiny multiple of a prime power; they ask whether the set of integers with this property has positive density or even density 1. The analogous question can be asked about those integers that cannot be the largest or second-largest denominator in an Egyptian fraction representation of 1, and so on (as is mentioned in [5]).

We can generalize this problem to Egyptian fraction representations of any positive rational number r. For a positive integer j, let us define

$$\mathcal{L}_j(r) = \{x \in \mathbb{Z}, \ x > r^{-1} : \text{there do not exist } x_1, \ldots, x_t \in \mathbb{Z},$$
$$x_1 > \ldots > x_t \geq 1 \text{ with } \sum_{i=1}^t 1/x_i = r \text{ and } x_j = x\},$$

the set of numbers that *cannot* be the jth-largest denominator in an Egyptian fraction representation of r. We exclude the integers $x \leq r^{-1}$ from consideration because they can never be a denominator in an Egyptian fraction representation of r (except for the trivial representation when r is itself the reciprocal of an integer). The questions of Erdős and Graham then become whether $\mathcal{L}_1(1)$ has positive density, what can be said about $\mathcal{L}_1(1) \cap \mathcal{L}_2(1)$, and so on.

In our primary theorem concerning these questions, we discover some information about the sets $\mathcal{L}_j(r)$ for $j \geq 2$ that is perhaps quite surprising:

THEOREM 2. *Let r be a positive rational number. The set $\mathcal{L}_j(r)$ is finite for any integer $j \geq 2$, and there exists an integer $j_0(r)$ such that $\mathcal{L}_j(r)$ is empty for all $j \geq j_0(r)$.*

For example, a consequence of Theorem 2 is that only finitely many numbers cannot be the second-largest denominator in an Egyptian fraction representation of 1; possibly $\{2, 4\}$ is a complete list of integers (greater than 1) with this property. Another consequence is that every integer greater than 1 can be the jth-largest denominator in an Egyptian fraction representation of 1, when j is sufficiently large; possibly this holds for every $j \geq 3$. It might be interesting to modify the proof of Theorem 2 to include explicit constants and subsequently determine the sets $\mathcal{L}_2(1)$ and $\mathcal{L}_3(1)$ precisely with the aid of a computer; however, we do not undertake these tasks herein.

Because the prime factors of the denominator of r are the only primes that can possibly be the largest denominator in an Egyptian fraction representation of r, the set $\mathcal{L}_1(r)$ is certainly infinite. However, we are able to answer Erdős and Graham's question of whether $\mathcal{L}_1(r)$ has positive density in the negative; in fact, we can even establish the order of growth of $\mathcal{L}_1(r)$. Let us define the counting function $L_1(r; x)$ of $\mathcal{L}_1(r)$,

$$L_1(r; x) = \#\{1 \leq n \leq x : n \in \mathcal{L}_1(r)\}.$$

Then we have the following theorem:

THEOREM 3. *Let r be a positive rational number. The set $\mathcal{L}_1(r)$ has zero density, and in fact, if $x \geq 3$ is a real number then*

$$(4) \qquad \frac{x \log \log x}{\log x} \ll_r L_1(r; x) \ll_r \frac{x \log \log x}{\log x}.$$

The lower bound in the inequality (4) is a simple quantitative consequence of the observation that tiny multiples of prime powers are elements of $\mathcal{L}_1(r)$, as we shall show in the next section, while the upper bound reflects the discovery that all elements of $\mathcal{L}_1(r)$ are of this form (the only ambiguity being the exact meaning of "tiny"). This discovery was probably known to Croot (certainly in the case where r is an integer), since the methods we shall use to establish the inequality (4) are to a large extent present in [3].

We mention that Theorem 1 has the following corollary:

Let r be a positive rational number. For every x that is sufficiently large in terms of r, there is a set $\mathcal{E}$ of integers not exceeding x, such that $\sum_{n \in \mathcal{E}} 1/n = r$ and

$$(5) \qquad |\mathcal{E}| > (1 - e^{-r})x - O_r\left(\frac{x \log \log x}{\log x}\right).$$

This corollary can also be seen to be best possible. This corollary improves the author's theorem from [8] (which explains the title of the current

paper), and in fact it was in the course of establishing this result that the methods of proof of the above theorems were developed. Here it is appropriate to note that recently, Croot [2] established a related theorem that any positive rational r has an Egyptian fraction representation whose denominators all lie in the interval $[y, (e^r + o(1))y]$, if y is large enough in terms of r; his result also implies the corollary just stated.

2. Reduction of Theorem 1. We begin by defining some notation that will be used throughout the paper. Hereafter p will always denote a prime and q will always denote a (not necessarily proper) prime power. As is standard, the function $\pi(x)$ denotes the number of primes not exceeding x, and $P(n)$ denotes the largest prime divisor of n. It is more convenient for our purposes, however, to regard the sequence of prime powers as more fundamental than the sequence of primes. Therefore, we shall use $\pi^*(x)$ to denote the number of prime powers not exceeding x, so that $\pi^*(x) = \pi(x) + \pi(x^{1/2}) + \pi(x^{1/3}) + \ldots$ We will also let $P^*(n)$ denote the largest prime power that divides n; for example, $P(12) = 3$ but $P^*(12) = 4$. (By convention we set $P^*(1) = 1$.) Notice that if n is the least common multiple of l and m, then $P^*(n) = \max\{P^*(l), P^*(m)\}$; in particular, if a and b are coprime and $a/b = a_1/b_1 + a_2/b_2$, then $P^*(b) \leq \max\{P^*(b_1), P^*(b_2)\}$.

Theorem 1 can be reduced to the following two propositions:

PROPOSITION 4. *Let I be a closed subinterval of $(0, \infty)$. There exists a positive real number $T(I)$ such that, for all integers $t > T(I)$ and all rational numbers $r = a/b \in I$ such that $P^*(b) < t \log^{-22} t$, there is a set $\mathcal{E}$ of t distinct positive integers such that $\sum_{n \in \mathcal{E}} 1/n = r$ and*

$$(6) \qquad \max\{n \in \mathcal{E}\} < \frac{t}{1 - e^{-r}} + O_I\left(\frac{t \log \log t}{\log t}\right).$$

PROPOSITION 5. *Let r be a positive rational number. There exists a positive constant $\delta(r)$ such that, for every real number x that is sufficiently large in terms of r, all sets $\mathcal{E}$ of positive integers not exceeding x for which $\sum_{n \in \mathcal{E}} 1/n = r$ satisfy*

$$|\mathcal{E}| \leq (1 - e^{-r})x - \delta(r)\frac{x \log \log x}{\log x}.$$

Let us verify that Theorem 1 follows from Propositions 4 and 5. Given a positive rational number $r = a/b$, define I to be the length-zero interval $I = \{r\}$, and choose an integer $t > T(I)$ so large that $t \log^{-22} t > P^*(b)$. We can apply Proposition 4 to obtain a set $\mathcal{E}$ of t distinct positive integers such that $\sum_{n \in \mathcal{E}} 1/n = r$, whose largest term is at most $t/(1 - e^{-r}) + O_r(t \log \log t / \log t)$. By the definition (1) of $M_t(r)$, this establishes the upper bound implicit in the asymptotic formula (3) when t is sufficiently large in terms of r; but by adjusting the constant implicit in the O-notation if necessary, we see that this upper bound is valid for all $t \geq t_0(r)$.

On the other hand, if $\mathcal{E}$ is any set of t distinct positive integers satisfying $\sum_{n \in \mathcal{E}} 1/n = r$ whose largest element is x_1, then Proposition 5 shows that

$$t \leq (1 - e^{-r})x_1 - \delta(r)\frac{x_1 \log \log x_1}{\log x_1},$$

which implies that

$$(7) \qquad x_1 \geq \frac{t}{1 - e^{-r}} + \delta(r)\frac{t \log \log t}{\log t}$$

when t is large enough in terms of r. Since $M_t(r)$ equals the smallest such x_1, we see that the right-hand side of the inequality (7) is also a lower bound for $M_t(r)$. This argument shows that Theorem 1 follows in its entirety from the two propositions.

The reader will have noticed that although Proposition 4 is stated uniformly for certain rational numbers in the interval I, no use was made of this in deducing Theorem 1. However, we shall need the uniformity present in Proposition 4 in the proofs of Theorems 2 and 3; it is for this reason that we take the time to establish the proposition in its current form.

It turns out that the construction used to establish Proposition 4 proceeds in two stages which, although similar in spirit, require quite different subsidiary lemmas to complete. For this reason, we reduce Proposition 4 to the following two propositions:

PROPOSITION 6. *Let I be a closed subinterval of $(0, \infty)$. For any real number x that is sufficiently large in terms of I, any rational number $r \in I$ whose denominator is not divisible by any prime power exceeding $x \log^{-22} x$, and any integer R satisfying*

$$(8) \qquad \left| (1 - e^{-r})x - \left(22(1 - e^{-r}) - \frac{3r}{e^r} \right)\frac{x \log \log x}{\log x} - R \right| < \frac{r}{e^r} \cdot \frac{x \log \log x}{\log x},$$

there exists a set $\mathcal{R}$ of integers satisfying:

(i) $\mathcal{R}$ *is contained in* $[x/(2e^r), x]$;
(ii) $|\mathcal{R}| = R$;

(iii) *if $a/b = r - \sum_{n \in \mathcal{R}} 1/n$ in lowest terms, then $1/\log x < a/b < 1$ and* $P^*(b) \le x^{1/5}$.

(As one might think, the constants in the inequality (8), other than the initial $(1 - e^{-r})$, are somewhat arbitrary and chosen simply for convenience during the proof.)

PROPOSITION 7. *Let y be a sufficiently large real number, and let a/b be a rational number satisfying $1/\log y < a/b < 1$ and $P^*(b) \le y$. Then there is a set $\mathcal{S}$ of integers satisfying*:

   (i) $\mathcal{S}$ *is contained in* $[1, 2y^4]$;
   (ii) $|\mathcal{S}| = 2\pi^*(y)$;
   (iii) $a/b = \sum_{n \in \mathcal{S}} 1/n$.

To see how Propositions 6 and 7 together imply Proposition 4, we fix a closed interval $I \subset (0, \infty)$, an integer $t > T(I)$ where $T(I)$ is a positive constant that is sufficiently large in terms of I, and a rational number $r \in I$ whose denominator is not divisible by any prime power exceeding $t \log^{-22} t$. We define

$$(9) \qquad x = \frac{t}{1 - e^{-r}} + \left(\frac{22 - (22 + 3r)e^{-r}}{(1 - e^{-r})^2} \right) \frac{t \log \log t}{\log t}$$

and $R = t - 2\pi^*(x^{1/5})$. When these values of x and R are substituted into (8), the left-hand side has order of magnitude $t(\log \log t)^2/\log^2 t$ after simplification, while the right-hand side has order of magnitude $t \log \log t/\log t$; therefore the inequality (8) holds as long as $T(I)$ is large enough. Certainly $x \log^{-22} x \ge t \log^{-22} t$ as well if $T(I)$ is large enough. We may therefore apply Proposition 6 to obtain a set $\mathcal{R}$ of integers and a rational number a/b satisfying properties (i)–(iii) of Proposition 6. With this rational a/b, we may then apply Proposition 7 with $y = x^{1/5}$ as long as $T(I)$ is large enough (since x and y are functions of t), obtaining a set $\mathcal{S}$ satisfying properties (i)–(iii) of that proposition.

We now set $\mathcal{E} = \mathcal{R} \cup \mathcal{S}$. Because $2y^4 = 2x^{4/5} < x/(2e^r)$ if $T(I)$ is large enough, it follows from the two properties (i) that $\mathcal{R}$ and $\mathcal{S}$ are disjoint and that the integers in $\mathcal{E}$ do not exceed x. Moreover, the two properties (ii) imply that $|\mathcal{E}| = t$, while the two properties (iii) imply that $r = \sum_{n \in \mathcal{E}} 1/n$. By the definition (9) of x, we see that this set $\mathcal{E}$ satisfies all the properties required for the conclusion of Proposition 4, and so that proposition does indeed follow from Propositions 6 and 7.

In a sense, we have separated the desire to have an Egyptian fraction with many terms from the desire to have a specific number of unit fractions that add to r. Proposition 6 yields an Egyptian fraction with many terms, but one whose sum is only an approximation to r; while Proposition 7 yields an exact Egyptian fraction representation of a/b with a specified number of

terms, but with that number of terms rather small compared to the size of the largest denominator.

The broad idea of the proofs of these two propositions uses the general strategy employed by the author in [8] and by Croot in [3]: a collection of unit fractions whose sum is relatively close to the target rational number is constructed, and then for each prime power that appears in the denominator of this sum but not in the denominator of the target rational, a few terms are omitted or added so that the modified sum is no longer divisible by that prime power. When all of the unwanted prime powers have been eradicated in this way, estimates on the number and sizes of the omitted or added terms are used to show that the resulting sum must exactly equal the target rational. Proposition 6 is used to evict the larger unwanted prime powers, while Proposition 7 is used to evict the smaller prime powers.

To summarize the achievements of this section, we have reduced Theorem 1 to establishing Propositions 5, 6, and 7. These three propositions will be the subjects of Sections 3, 4, and 5, respectively. To complete the outline of the rest of this paper, we mention that Theorem 2 will be established in Section 6 and Theorem 3 will be established in Section 7.

3. The very large prime powers. In this section we establish Proposition 5, which was used to show that Theorem 1 is best possible. The strategy is to quantify the observation that a tiny multiple of a very large prime (or prime power) cannot appear in an Egyptian fraction representation of a given rational number, and then to calculate the effect that this restriction has on the possible number of terms in such a representation.

LEMMA 8. *Let x be a positive real number. Suppose that $x_1, \ldots, x_t$ are distinct positive integers not exceeding x and that p is a prime dividing at least one of the x_i. If p does not divide the denominator of $\sum_{i=1}^{t} 1/x_i$ expressed in lowest terms, then $p \ll x/\log x$.*

We remark that a slightly modified version of this lemma could be established for prime powers q rather than merely for primes p, but this formulation suffices for our purposes.

P r o o f. Because the integers x_i that are not divisible by p do not affect whether the denominator of $\sum_{i=1}^{t} 1/x_i$ is divisible by p, we may assume that p divides all of the x_i. Set $w_i = x_i/p$ for each $1 \le i \le t$, and set $\lambda = \mathrm{lcm}\{w_1, \ldots, w_t\}$. Then

$$(10) \qquad \sum_{i=1}^{t} \frac{1}{x_i} = \left(\sum_{i=1}^{t} \frac{\lambda p}{x_i} \right) \Big/ (\lambda p),$$

where each summand is an integer by the definition of λ. We are assuming that the denominator of $\sum_{i=1}^{t} 1/x_i$ is not divisible by p when reduced to

lowest terms. Consequently, if

$$(11) \qquad N = \sum_{i=1}^{t} \frac{\lambda p}{x_i} = \lambda \sum_{i=1}^{t} \frac{1}{w_i}$$

is the numerator of the fraction on the right-hand side of equation (10), then N must be a positive multiple of p; in particular, $N \geq p$ and thus $\log N \geq \log p$.

On the other hand, the collection $\{w_i\}$ is a subset of the integers not exceeding x/p. If we define $L(x) = \mathrm{lcm}\{1, 2, \ldots, \lfloor x \rfloor\}$, so that

$$(12) \qquad L(x) = \prod_{p^\nu \leq x} p = \exp \Big(\sum_{p^\nu \leq x} \log p \Big) \leq e^{2x}$$

by the prime number estimate of Chebyshev, then $\lambda \leq L(x/p)$ and so equation (11) implies

$$N \leq L\Big(\frac{x}{p}\Big) \sum_{w \leq x/p} \frac{1}{w} \leq L\Big(\frac{x}{p}\Big)\Big(\log \frac{x}{p} + 1\Big).$$

But then

$$\log p \leq \log N \leq \log L\Big(\frac{x}{p}\Big) + O\Big(\log\log\frac{x}{p}\Big) \ll \frac{x}{p}$$

by the estimate (12), and so $p \ll x/\log x$ as claimed. $\blacksquare$

Next we establish an elementary lemma that provides asymptotic formulae for the number of integers free of very large prime (or prime power) factors and for the sum of the reciprocals of such integers.

LEMMA 9. *Uniformly for $\sqrt{x} \leq y \leq x$ and $0 < \alpha < 1$, we have*

$$\sum_{\substack{\alpha x \leq n \leq x \\ P(n) > y}} 1 = (1 - \alpha)x \log \frac{\log x}{\log y} + O\Big(\frac{x}{\log x}\Big)$$

and

$$\sum_{\substack{\alpha x \leq n \leq x \\ P(n) > y}} \frac{1}{n} = \log \alpha^{-1} \log \frac{\log x}{\log y} + O\Big(\frac{1}{\alpha \log x}\Big).$$

Both formulas remain valid if $P(n)$ is replaced by $P^*(n)$ in the conditions of summation.

P r o o f. It is a direct consequence of Mertens' formula for $\sum_{p \leq x} 1/p$ that

$$(13) \qquad \sum_{y < p \leq x} \frac{1}{p} = \log \frac{\log x}{\log y} + O\Big(\frac{1}{\log y}\Big),$$

and the same asymptotic formula is true if the summation is taken over prime powers q rather than merely primes p. We note that any integer $n \leq x$ such that $P(n) > y$ can be written as $n = mp$ where $y < p \leq x$, and this representation is unique since $y \geq \sqrt{x}$. The first assertion of the lemma then follows from equation (13) by writing

$$\sum_{\substack{\alpha x \leq n \leq x \\ P(n) > y}} 1 = \sum_{y < p \leq x} \sum_{\alpha x/p \leq m \leq x/p} 1 = \sum_{y < p \leq x} \left(\frac{(1 - \alpha)x}{p} + O(1) \right)$$

and invoking Chebyshev's estimate $\pi(x) \ll x/\log x$ to bound the error term. The second assertion follows in a similar manner from writing

$$\sum_{\substack{\alpha x \leq n \leq x \\ P(n) > y}} \frac{1}{n} = \sum_{y < p \leq x} \frac{1}{p} \sum_{\alpha x/p \leq m \leq x/p} \frac{1}{m} = \sum_{y < p \leq x} \frac{1}{p} \left(\log \alpha^{-1} + O\left(\frac{p}{\alpha x} \right) \right).$$

Because the asymptotic formula (13) is insensitive to the inclusion of the proper prime powers, these arguments are equally valid when $P(n)$ is replaced by $P^*(n)$ in the conditions of summation. ∎

Proof of Proposition 5. Suppose that r is a positive rational number, x is a real number that is sufficiently large in terms of r, and $\mathcal{E}$ is a set of positive integers not exceeding x such that $\sum_{n \in \mathcal{E}} 1/n = r$. Let $C > 1$ be a large constant and define

$$\mathcal{A} = \{n \leq x : P(n) \leq Cx/\log x\}.$$

We can assume that x is so large that all of the prime divisors of the denominator of r are less than $Cx/\log x$. Then if C is chosen large enough, the set $\mathcal{E}$ must be contained in $\mathcal{A}$ by Lemma 8.

Choose $0 < \alpha < 1$ such that, if we set $\mathcal{E}' = [\alpha x, x] \cap \mathcal{A}$, then $|\mathcal{E}'| = |\mathcal{E}|$; in other words, $\mathcal{E}'$ is the subset of $\mathcal{A}$ with cardinality $|\mathcal{E}|$ whose elements are as large as possible. Then

$$|\mathcal{E}| = |\mathcal{E}'| = \sum_{\alpha x \leq n \leq x} 1 - \sum_{\substack{\alpha x \leq n \leq x \\ P(n) > Cx/\log x}} 1$$

$$= (1 - \alpha)x + O(1) - (1 - \alpha)x \log \left(\frac{\log x}{\log(Cx/\log x)} \right) + O\left(\frac{x}{\log x} \right)$$

by Lemma 9 with $y = Cx/\log x$. Since

$$(14) \qquad \log \left(\frac{\log x}{\log(Cx/\log x)} \right) = \log \left(1 - \frac{\log \log x}{\log x} + \frac{\log C}{\log x} \right)^{-1}$$

$$= \frac{\log \log x}{\log x} + O\left(\frac{1}{\log x} \right),$$

we see that

$$(15) \qquad |\mathcal{E}| = (1-\alpha)x\left(1 - \frac{\log\log x}{\log x}\right) + O\left(\frac{x}{\log x}\right).$$

On the other hand, the elements of $\mathcal{E}'$ are by definition at least as big as the elements of $\mathcal{E}$, and so

$$r = \sum_{n\in\mathcal{E}}\frac{1}{n} \geq \sum_{n\in\mathcal{E}'}\frac{1}{n} = \sum_{\alpha x\leq n\leq x}\frac{1}{n} - \sum_{\substack{\alpha x\leq n\leq x \\ P(n)>Cx/\log x}}\frac{1}{n}$$

$$= \log\alpha^{-1} + O\left(\frac{1}{\alpha x}\right) - \log\alpha^{-1}\log\frac{\log x}{\log(Cx/\log x)} + O\left(\frac{1}{\alpha\log x}\right),$$

again by Lemma 9 with $y = Cx/\log x$. Using equation (14) again, we see that

$$r \geq \log\alpha^{-1}\left(1 - \frac{\log\log x}{\log x}\right) + O\left(\frac{1}{\alpha\log x}\right),$$

which implies that

$$\alpha \geq e^{-r}\left(1 - \frac{r\log\log x}{\log x} + O_r\left(\frac{1}{\log x}\right)\right).$$

With this lower bound, equation (15) becomes

$$|\mathcal{E}| \leq \left(1 - e^{-r}\left(1 - \frac{r\log\log x}{\log x} + O_r\left(\frac{1}{\log x}\right)\right)\right)x\left(1 - \frac{\log\log x}{\log x}\right)$$
$$+ O\left(\frac{x}{\log x}\right)$$

$$= (1 - e^{-r})x - (1 - e^{-r}(1+r))\frac{x\log\log x}{\log x} + O_r\left(\frac{x}{\log x}\right).$$

Since $e^r > 1 + r$ for $r > 0$, we may choose $\delta(r)$ satisfying $0 < \delta(r) < 1 - e^{-r}(1+r)$, whence

$$(16) \qquad |\mathcal{E}| \leq (1 - e^{-r})x - \delta(r)\frac{x\log\log x}{\log x}$$

when x is large enough in terms of r. ∎

4. The large prime powers. In this section we establish Proposition 6. The methods in this section are in large part derived from those in Croot [3], albeit with some modifications necessary for the problem at hand; in particular, our Lemma 11 below is a direct generalization of [3, Proposition 2].

Let $\|x\|$ denote the distance from x to the nearest integer. If a and n are coprime, then let $\bar{a}$ (mod n) denote the integer b with $0 < b < n$ and $ab \equiv 1$ (mod n). (Often we write simply $\bar{a}$ when the modulus is clear from the context, e.g., when the term $\bar{a}$ appears in the numerator of a

fraction whose denominator is the modulus.) The following lemma demonstrates that under suitable conditions on a set $\mathcal{M}$ of integers, the elements of $\mathcal{M}$ cannot all be small compared to a modulus n and yet predominantly have inverses (mod n) that are close to 0 (mod n), even when scaled by a nonzero residue h.

LEMMA 10. *Let n be a sufficiently large integer, let k be a positive integer, and let B and C be positive numbers with C satisfying $200(\log n/\log \log n)^k < C < n$. Suppose that $\mathcal{M}$ is a set of positive integers with cardinality greater than C, such that each element m of $\mathcal{M}$ is less than B and is the product of k distinct primes not dividing n. Then for any $0 < h < n$, at least $C/2$ elements m of $\mathcal{M}$ satisfy*

$$\left\| \frac{h\overline{m}}{n} \right\| > \frac{C(\log \log n)^k}{200B \log^k n},$$

where $\overline{m}$ denotes the inverse of m (mod n).

P r o o f. For $m \in \mathcal{M}$, define r_m to be the integer satisfying $-n/2 < r_m \leq n/2$ and $r_m \equiv h\overline{m}$ (mod n); since n does not divide h and $(m,n) = 1$, we see that r_m is nonzero. Also define

$$s_m = \frac{mr_m - h}{n},$$

so that s_m is an integer satisfying $|s_m| < m/2 + 1$. Suppose that at least $C/2$ of the s_m satisfied $|s_m| < C(\log \log n)^k/(100 \log^k n)$. Then, by the pigeonhole principle, there would be an s with $|s| < C(\log \log n)^k/(100 \log^k n)$ such that $s_m = s$ for at least

$$\frac{C/2}{2C(\log \log n)^k/(100 \log^k n) + 1} > 20 \left(\frac{\log n}{\log \log n} \right)^k$$

of the elements m of $\mathcal{M}$, by the lower bound on C. For each such m, we see that

$$r_m = \frac{ns_m + h}{m} = \frac{ns + h}{m};$$

and since the r_m are nonzero integers, we see that the nonzero integer $ns+h$ is divisible by at least $20(\log n/\log \log n)^k$ elements of $\mathcal{M}$.

On the other hand, it is well known that the maximal order of the number of distinct prime divisors of an integer m is asymptotic to $\log m/\log \log m$, as achieved by those m that are the product of all the primes up to about $\log m$; thus when m is sufficiently large, every integer up to m has less than $2 \log m/\log \log m$ distinct prime divisors. Since

$$|ns + h| < n(C(\log \log n)^k/(100 \log^k n) + 1) < n^2$$

by the upper bound on C, we see that $ns+h$ has at most $2 \log n^2/\log \log n^2 < 4 \log n/\log \log n$ distinct prime factors when n is sufficiently large, and so the

number of divisors of $ns + h$ that are the product of k distinct primes does not exceed

$$\frac{1}{k!}\left(\frac{4\log n}{\log\log n}\right)^k < 20\left(\frac{\log n}{\log\log n}\right)^k,$$

contradicting the lower bound for the number of elements of $\mathcal{M}$ that divide $ns + h$.

This contradiction shows that at most $C/2$ of the s_m satisfy $|s_m| < C(\log\log n)^k/(100\log^k n)$, and so at least $C/2$ of the s_m satisfy the reverse inequality. For these elements m, we see that

$$|r_m| = \frac{|ns_m + h|}{m} > \frac{n}{m}\left(\frac{C(\log\log n)^k}{100\log^k n} - 1\right) > \frac{Cn(\log\log n)^k}{200B\log^k n}$$

by the upper bounds for h and m and the lower bound for C. But then by the definition of r_m,

$$\left\|\frac{h\overline{m}}{n}\right\| = \left\|\frac{r_m}{n}\right\| > \frac{C(\log\log n)^k}{200B\log^k n},$$

which establishes the lemma. ∎

The next lemma translates the statement of Lemma 10 to an assertion that under suitable conditions on a set $\mathcal{M}$, the inverses $(\bmod\ n)$ of the elements of $\mathcal{M}$ must have a subset whose sum is congruent to any predetermined residue class $(\bmod\ n)$.

LEMMA 11. *Let n be a sufficiently large integer, let k be a positive integer, and let B be a real number satisfying*

$$(17) \qquad\qquad B > \frac{(\log n)^{(k-1)/2}}{(\log\log n)^{k/2}}.$$

Suppose that $\mathcal{M}$ is a set of integers whose cardinality C satisfies

$$(18) \qquad\qquad C > \frac{200B^{2/3}(\log n)^{(2k+1)/3}}{(\log\log n)^{2k/3}},$$

such that each element m of $\mathcal{M}$ is less than B and is the product of k distinct primes not dividing n. Then for any residue class $a\ (\bmod\ n)$, there is a subset $\mathcal{K}$ of $\mathcal{M}$ such that

$$(19) \qquad\qquad \sum_{m\in\mathcal{K}} \overline{m} \equiv a\ (\bmod\ n).$$

P r o o f. We begin by remarking, in preparation for applying Lemma 10, that the hypotheses on B and C ensure that $C > 200(\log n/\log\log n)^k$. Also, if $C \geq n$, then the conclusion of the lemma holds under the weaker assumption that each element of $\mathcal{M}$ is coprime to n, by the Cauchy–Davenport–

Chowla Theorem (see for instance Vaughan [9, Lemma 2.14], and also [8, Lemma 2]). Therefore we can assume that $C < n$.

Let $e_n(x)$ denote the complex exponential $e^{2\pi i x/n}$ of period n. If we let N be the number of subsets of $\mathcal{M}$ satisfying the condition (19), then by the finite Fourier transform

$$
(20) \qquad N = \sum_{\mathcal{K} \subset \mathcal{M}} \frac{1}{n} \sum_{h=0}^{n-1} e_n\Big(h\Big(\sum_{m \in \mathcal{K}} \overline{m} - a\Big)\Big)
$$

$$
= \frac{1}{n} \sum_{h=0}^{n-1} e_n(-ha) \prod_{m \in \mathcal{M}} (1 + e_n(h\overline{m}))
$$

$$
= \frac{2^C}{n} + \frac{1}{n} \sum_{h=1}^{n-1} e_n(-ha) P_h,
$$

where $P_h = \prod_{m \in \mathcal{M}} (1 + e_n(h\overline{m}))$. Using the identity $|1 + e^{it}|^2 = 2 + 2\cos t$ and the inequality $1 + \cos 2\pi t \leq 2 - 8\|t\|^2$, we see that

$$
(21) \qquad |P_h|^2 = \prod_{m \in \mathcal{M}} \Big(2 + 2\cos\frac{2\pi h\overline{m}}{n}\Big) \leq 4^C \prod_{m \in \mathcal{M}} \Big(1 - 4\Big\|\frac{h\overline{m}}{n}\Big\|^2\Big).
$$

All of the terms in this product are nonnegative and bounded above by 1; and when n is sufficiently large, by Lemma 10 at least $C/2$ of them are bounded above by

$$
1 - 4\Big(\frac{C(\log\log n)^k}{200B\log^k n}\Big)^2
$$

when $1 \leq h \leq n - 1$. Using this fact in the inequality (21) along with the bound $1 - t \leq e^{-t}$, we obtain

$$
|P_h|^2 \leq 4^C \Big(1 - \frac{C^2(\log\log n)^{2k}}{10000B^2\log^{2k} n}\Big)^{C/2} \leq 4^C \exp\Big(\frac{-C^3(\log\log n)^{2k}}{20000B^2\log^{2k} n}\Big)
$$

$$
< 4^C \exp(-2\log n)
$$

by the lower bound (18) on C, and thus $|P_h| < 2^C/n$ when $1 \leq h \leq n - 1$.

From this upper bound, we deduce from equation (20) that

$$
\Big|N - \frac{2^C}{n}\Big| \leq \frac{1}{n} \sum_{h=1}^{n-1} |P_h| < \frac{(n-1)2^C}{n^2},
$$

which implies that $N > 2^C/n^2$. In particular, there do exist subsets $\mathcal{K}$ of $\mathcal{M}$ of the desired type. ∎

We remark that a stronger inequality for B than (17) might be needed to ensure the existence of a set $\mathcal{M}$ with the properties described in the statement of Lemma 11 (for instance, in the case $k = 1$).

The following lemma is the main tool that will be used in our recursive construction in the proof of Proposition 6.

LEMMA 12. *Let $0 < \xi < 1$ be a real number, and let x be a real number that is sufficiently large in terms of ξ. Let c/d be a rational number and define $q = P^*(d)$, and suppose that $x^{1/5} \le q \le x \log^{-22} x$. Then there exists a set $\mathcal{U}$ of integers satisfying*:

   (i) $\mathcal{U}$ *is contained in* $[\xi x, x]$;
  (ii) $|\mathcal{U}| \le 200(x/q)^{2/3} \log^3 x$;
 (iii) *for each element n of $\mathcal{U}$, $P^*(n) = q$*;
 (iv) *if $c'/d' = c/d - \sum_{n \in \mathcal{U}} 1/n$ in lowest terms, then $P^*(d') < q$*.

P r o o f. We apply Lemma 11 with $n = q$, $k = 4$, $B = x/q$, and a the residue class of $c\overline{(d/q)}$ (mod q). Let $\mathcal{P}$ be the set of all primes in the interval $((\xi x/q)^{1/4}, (x/q)^{1/4})$ that do not divide q, and let $\mathcal{M}_0$ be the set of all integers of the form $p_1 p_2 p_3 p_4$, where the p_i are distinct elements of $\mathcal{P}$. Certainly each element of $\mathcal{M}_0$ is between ξB and B. Since the cardinality of $\mathcal{P}$ is $\gg (1-\xi)(x/q)^{1/4} \log^{-1} x$ by the prime number theorem, the cardinality of $\mathcal{M}_0$ is $\gg_\xi x/(q \log^4 x)$, and so we can choose a subset $\mathcal{M}$ of cardinality

$$C = \lfloor 200(x/q)^{2/3} \log^3 x \rfloor$$

when x is sufficiently large in terms of ξ, by the upper bound on q.

Since these values of B and C do satisfy the hypotheses (17) and (18) of Lemma 11, we can find a subset $\mathcal{K}$ of $\mathcal{M}$ such that $\sum_{m \in \mathcal{K}} \overline{m} \equiv a$ (mod q). Now define

$$\mathcal{U} = \{qm : m \in \mathcal{K}\}.$$

We see immediately that properties (i) and (ii) hold for $\mathcal{U}$. Since $q \ge x^{1/5}$, all the elements of $\mathcal{P}$ (hence all prime-power divisors of elements of $\mathcal{K}$) are less than q, and so property (iii) holds for $\mathcal{U}$ as well. As for property (iv), it is clear that $P^*(d') \le q$, since all of the prime powers dividing d or any of the elements of $\mathcal{U}$ are at most q. On the other hand, q does not divide d', since if $q = p^\nu$ then

$$\frac{c'q}{d'} = \frac{cq}{d} - \sum_{n \in \mathcal{U}} \frac{q}{n} = \frac{c}{d/q} - \sum_{n \in \mathcal{M}} \frac{1}{m} \equiv c\overline{(d/q)} - \sum_{n \in \mathcal{M}} \overline{m} \equiv 0 \text{ (mod } p),$$

and so d' is divisible by at most $p^{\nu-1}$ after reducing to lowest terms. Thus $P^*(d') < q$, which establishes the lemma. ∎

As a last step in the construction used to establish Proposition 6, we shall be appending a collection of unit fractions none of whose denominators contain large prime-power divisors. The following lemma ensures that there are enough such integers in a suitable range to accommodate this.

LEMMA 13. *Let* $0 < \eta < 1$ *be a real number. There exists a constant* $\delta = \delta(\eta)$ *such that, for all real numbers* x *that are sufficiently large in terms of* η *and for all pairs of real numbers* $\eta \le \alpha, \varepsilon < 1$, *there are at least* δx *integers* n *in the interval* $[\alpha x/2, \alpha x]$ *satisfying* $P^*(n) \le x^\varepsilon$.

P r o o f. We recall that a y-smooth integer is one all of whose prime factors are at most y, so that in particular, an integer n is x^ε-smooth precisely when it satisfies $P(n) \le x^\varepsilon$. It is well-known (see for instance Hildebrand and Tenenbaum [6]) that the number of y-smooth integers $n \le x$ is

$$(22) \qquad x\varrho\left(\frac{\log x}{\log y}\right) + O_\eta\left(\frac{x}{\log x}\right)$$

uniformly for $x^\eta \le y \le x$; here $\varrho(u)$ is the Dickman function, which is positive, continuously differentiable on $[1, \infty)$, and satisfies $|\varrho'(u)| \le 1$ on that interval. In particular,

$$(23) \qquad \varrho\left(\frac{\log \alpha x}{\log x^\varepsilon}\right) = \varrho\left(\varepsilon^{-1} + \frac{\log \alpha}{\varepsilon \log x}\right) = \varrho(\varepsilon^{-1}) + O_\eta\left(\frac{1}{\log x}\right)$$

uniformly for $\eta \le \alpha, \varepsilon < 1$. However, n being x^ε-smooth is a slightly weaker condition than $P^*(n) \le x^\varepsilon$. The asymptotics for the counting function of x^ε-smooth integers could be shown to hold for the number of integers $n \le x$ satisfying $P^*(n) \le x^\varepsilon$ as well, but since we only need a weak lower bound for the number of such integers we argue as follows.

By (22), using equation (23) and the analogous statement with α replaced by $\alpha/2$, we see that the number of x^ε-smooth integers between $\alpha x/2$ and αx is $\alpha\varrho(\varepsilon^{-1})x/2 + O_\eta(x/\log x)$. On the other hand, the integers n that are x^ε-smooth but for which $P^*(n) > x^\varepsilon$ are all divisible by at least one prime power $p^\nu > x^\varepsilon$ with $p \le x^\varepsilon$. Given an integer $k \ge 3$, any integer n that is x^ε-smooth but for which $P^*(n) > x^\varepsilon$ must either be divisible by the kth power of some prime, or else by the square of some prime exceeding $x^{\varepsilon/k}$. The number of integers $\alpha x/2 \le n \le \alpha x$ that are divisible by the kth power of a prime is $\alpha(1 - \zeta(k)^{-1})x/2 + O(x^{1/k})$, while the number of integers $n \le x$ that are divisible by the square of a prime exceeding $x^{\varepsilon/k}$ is at most

$$\sum_{x^{\varepsilon/k} < p \le x^{1/2}} \frac{x}{p^2} + 1 \ll x^{1 - \varepsilon/k}.$$

Therefore, if we choose k so large that $\zeta(k)^{-1} > 1 - \varrho(\eta^{-1})$, then the number of integers $\alpha x/2 \le n \le \alpha x$ such that $P^*(n) \le x^\varepsilon$ is at least

$$\left(\frac{\alpha\varrho(\varepsilon^{-1})x}{2} + O_\eta\left(\frac{x}{\log x}\right)\right) - \left(\frac{\alpha(1 - \zeta(k)^{-1})x}{2} + O(x^{1/k})\right) - O(x^{1-\varepsilon/k}) > \delta x$$

for some constant $\delta = \delta(\eta)$, as long as x is sufficiently large in terms of η. ∎

Proof of Proposition 6. Let $I = [m, M]$ be a closed subinterval of $(0, \infty)$, and let x be a real number that is sufficiently large in terms of I. Let $r \in I$ be a rational number whose denominator is not divisible by any prime power exceeding $x \log^{-22} x$. Let

$$\alpha = e^{-r}, \quad \eta = \min\{e^{-M}, 1/5\}, \quad \xi = e^{-m},$$

so that $0 < \eta \leq \alpha \leq \xi < 1$. Since we are assuming that x is sufficiently large in terms of I, we can assume in particular that x is sufficiently large in terms of ξ and η when appealing to Lemmas 12 and 13. Define $\mathcal{A}$ to be the set of all integers n in $[\alpha x, x]$ such that $P^*(n) \leq x \log^{-22} x$, and set

$$z = \pi^*(x \log^{-22} x) \quad \text{and} \quad z' = \pi^*(x^{1/5}).$$

Let $\{q_1, q_2, \ldots\}$ denote the sequence of prime powers in increasing order, and let p_i denote the prime of which q_i is a power. (By convention we set $p_0 = q_0 = 1$.)

Our strategy is to recursively define a sequence $\{a_i/b_i\}$ $(z+1 \geq i \geq z')$ of rationals that increase in size as the index i decreases, such that the largest prime-power divisor of each b_i is less than q_i. The first member a_{z+1}/b_{z+1} will be the difference between our original r and the sum of the reciprocals of the elements of $\mathcal{A}$, and each a_i/b_i will be obtained from the previous a_{i+1}/b_{i+1} by adding several unit fractions whose denominators belong to $\mathcal{A}$. The collection of all elements of $\mathcal{A}$ not involved in this construction will almost be the set $\mathcal{R}$ described in Proposition 6. This collection will have slightly fewer than the desired R elements, but we shall rectify the error simply by appending the appropriate number of integers without large prime-power factors from the interval $[\alpha x/2, \alpha x]$, and the resulting collection will be our set $\mathcal{R}$.

Define

$$\frac{a_{z+1}}{b_{z+1}} = r - \sum_{n \in \mathcal{A}} \frac{1}{n}.$$

We recursively define rationals $\{a_z/b_z, a_{z-1}/b_{z-1}, \ldots, a_{z'}/b_{z'}\}$ and sets $\{\mathcal{R}_z, \ldots, \mathcal{R}_{z'}\}$ of integers as follows. Suppose first that q_i divides b_i; then we apply Lemma 12 with $c/d = a_{i+1}/b_{i+1}$ and $q = q_i$. The lemma requires that $q_i = P^*(b_{i+1})$, and since we are supposing that q_i divides b_i we need only ensure that $P^*(b_{i+1}) \leq q_i$. In the case $i = z$, the inequality $P^*(b_{i+1}) \leq q_i$ is equivalent to $P^*(b_{z+1}) \leq x \log^{-22} x$, which is satisfied by the definitions of a_{z+1}/b_{z+1} and $\mathcal{A}$ and the hypothesis that the denominator of r is not divisible by any prime power exceeding $x \log^{-22} x$. On the other hand, the inequality $P^*(b_{i+1}) \leq q_i$ will be satisfied for smaller values of i by the recursive construction (as we shall see in a moment). Let $\mathcal{R}_i$ be the set $\mathcal{U}$

obtained from applying Lemma 12, and let

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

in lowest terms, so that by the lemma, $P^*(b_i) < q_i$ and thus $P^*(b_i) \leq q_{i-1}$ (justifying the claim of the previous sentence).

On the other hand, if q_i does not divide b_i, then we simply set $\mathcal{R}_i = \emptyset$ and $a_i/b_i = a_{i+1}/b_{i+1}$; since $P^*(b_{i+1}) \leq q_i$ by the recursive construction and q_i does not divide b_{i+1}, we see that $P^*(b_i) \leq q_{i-1}$.

Notice that each $\mathcal{R}_i$ is a subset of $\mathcal{A}$, since $\mathcal{R}_i$ is either empty or else (by Lemma 12) is contained in $[\xi x, x] \subset [\alpha x, x]$, and each element n of $\mathcal{R}_i$ satisfies $P^*(n) = q_i \leq x \log^{-22} x$. Notice also that the various $\mathcal{R}_i$ are pairwise disjoint, again since $P^*(n) = q_i$ for $n \in \mathcal{R}_i$.

Now set $\mathcal{R}' = \mathcal{A} \setminus \bigcup_{i=z'}^{z} \mathcal{R}_i$, so that

$$r = \frac{a_{z'}}{b_{z'}} + \sum_{n \in \mathcal{R}'} \frac{1}{n}$$

(here we have used the disjointness of the $\mathcal{R}_i$). We note that the cardinality of $\mathcal{A}$ is

$$(24) \quad |\mathcal{A}| = \sum_{\substack{\alpha x \leq n \leq x \\ P^*(n) \leq x \log^{-22} x}} 1 = (1 - \alpha)x + O(1) - \sum_{\substack{\alpha x \leq n \leq x \\ P^*(n) > x \log^{-22} x}} \frac{1}{n}$$

$$= (1 - \alpha)x - \frac{22(1 - \alpha)x \log \log x}{\log x} + O\left(\frac{x}{\log x}\right)$$

by applying Lemma 9 with $y = x \log^{-22} x$. On the other hand, $\mathcal{R}'$ is a subset of $\mathcal{A}$, while the set $\mathcal{A} \setminus \mathcal{R}'$ is simply the union of all the $\mathcal{R}_i$ and thus has cardinality

$$\left| \bigcup_{z' \leq i \leq z} \mathcal{R}_i \right| = \sum_{z' \leq i \leq z} |\mathcal{R}_i| \leq \sum_{x^{1/5} \leq q \leq x \log^{-22} x} 200 \left(\frac{x}{q}\right)^{2/3} \log^3 x \ll x \log^{-4} x$$

by Mertens' formula (13) and partial summation. Therefore the last expression in equation (24) represents the cardinality of $\mathcal{R}'$ as well.

By the hypothesis (8) on R, we see that R exceeds the cardinality of $\mathcal{R}'$, but by no more than $(4\alpha \log \alpha^{-1})x \log \log x / \log x$. Let $\mathcal{R}''$ be any set of $R - |\mathcal{R}'|$ integers n from the interval $[\alpha x/2, \alpha x]$, each satisfying $P^*(n) \leq x^{1/5}$; we can find such a set by Lemma 13 with $\varepsilon = 1/5$, since

$$(4\alpha \log \alpha^{-1}) \frac{x \log \log x}{\log x} \leq \frac{4x \log \log x}{e \log x} < \delta x$$

when x is sufficiently large in terms of $\delta = \delta(\eta) = \delta(I)$.

Now set $\mathcal{R} = \mathcal{R}' \cup \mathcal{R}''$, so that $\mathcal{R}$ is contained in the interval $[\alpha x/2, x] = [x/(2e^r), x]$ and the cardinality of $\mathcal{R}$ is precisely R, and set

$$\frac{a}{b} = r - \sum_{n \in \mathcal{R}} \frac{1}{n} = \frac{a_{z'}}{b_{z'}} - \sum_{n \in \mathcal{R}''} \frac{1}{n}.$$

Since $P^*(b_{z'}) < q'_z \le x^{1/5}$ by the definition of z', and $P^*(n) \le x^{1/5}$ for all $n \in \mathcal{R}''$ by the definition of $\mathcal{R}''$, we see that $P^*(b) \le x^{1/5}$. It only remains to show that $1/\log x < a/b < 1$ to establish the proposition.

We have

(25) $\qquad \dfrac{a}{b} = r - \displaystyle\sum_{n \in \mathcal{R}} \frac{1}{n} = \log \alpha^{-1} - \sum_{n \in \mathcal{A}} \frac{1}{n} + \sum_{z' \le i \le z} \sum_{n \in \mathcal{R}_i} \frac{1}{n} - \sum_{n \in \mathcal{R}''} \frac{1}{n}.$

From the definition of $\mathcal{A}$, and by Lemma 9 with $y = x \log^{-22} x$, the first sum in the last expression of equation (25) is

$$\sum_{\alpha x \le n \le x} \frac{1}{n} - \sum_{\substack{\alpha x \le n \le x \\ P^*(n) > x \log^{-22} x}} \frac{1}{n}$$

$$= \log \alpha^{-1} + O\left(\frac{1}{\alpha x}\right) - \frac{22 \log \alpha^{-1} \log \log x}{\log x} + O\left(\frac{1}{\alpha \log x}\right).$$

The double sum in equation (25) is at most

$$\sum_{x^{1/5} \le q \le x \log^{-22} x} \left(200 \left(\frac{x}{q}\right)^{2/3} \log^3 x\right) \frac{1}{\alpha x} \ll \frac{1}{\alpha \log^4 x},$$

and the last sum in equation (25) is nonnegative and at most

$$(R - |\mathcal{R}'|) \frac{2}{\alpha x} \le \frac{(4\alpha \log \alpha^{-1}) x \log \log x}{\log x} \cdot \frac{2}{\alpha x} = \frac{8 \log \alpha^{-1} \log \log x}{\log x}.$$

Consequently, equation (25) implies the inequalities

$$\frac{14 \log \alpha^{-1} \log \log x}{\log x} + O\left(\frac{1}{\eta \log x}\right) \le \frac{a}{b} \le \frac{22 \log \alpha^{-1} \log \log x}{\log x} + O\left(\frac{1}{\eta \log x}\right)$$

(since $\alpha \ge \eta$), which certainly implies that $1/\log x < a/b < 1$ when x is sufficiently large in terms of η. This establishes Proposition 6. ∎

5. The small prime powers. In this section we establish Proposition 7. We are now concerned more with having precise control over the number of terms in our Egyptian fractions than with sharply bounding the sizes of their denominators, as opposed to the case when we considered Proposition 6 in the previous section. The lemmas in this section appear in Croot [3], but we provide proofs for the sake of completeness and because we state the lemmas in somewhat different forms.

LEMMA 14. *Let $p^\nu \geq 5$ be a power of an odd prime p, and let a be any integer. There exist integers m_1 and m_2, satisfying $(p^\nu - 3)/2 \leq m_1 < m_2 < p^\nu$ and $p \nmid m_1 m_2$, such that $\overline{m}_1 + \overline{m}_2 \equiv a \pmod{p}$.*

P r o o f. Assume first that $p \geq 5$. Consider the set $\mathcal{M} = \{p^\nu - (p+3)/2 \leq m < p^\nu\}$ of $(p+3)/2$ integers, none of which is a multiple of p. Define two sets of residues $(\operatorname{mod} p)$:

$$\mathcal{M}_1 = \{\overline{m} : m \in \mathcal{M}\} \quad \text{and} \quad \mathcal{M}_2 = \{a - \overline{m} : m \in \mathcal{M}\},$$

where $\overline{m}$ denotes the multiplicative inverse of $m \pmod{p}$. Both $\mathcal{M}_1$ and $\mathcal{M}_2$ have $(p+3)/2$ distinct elements $(\operatorname{mod} p)$, and each $\mathcal{M}_i$ is a subset of the p residue classes $(\operatorname{mod} p)$, so by the pigeonhole principle there must be at least three residue classes m common to $\mathcal{M}_1$ and $\mathcal{M}_2$.

For any such m, if we let $m_1 = \overline{m}$ and $m_2 = \overline{a - m}$, then each m_i is in $\mathcal{M}$ by the definitions of the $\mathcal{M}_i$, and $\overline{m}_1 + \overline{m}_2 \equiv a \pmod{p}$. Furthermore, there is precisely one $m \pmod{p}$, namely $m \equiv \overline{2}a \pmod{p}$, such that $m_1 \equiv m_2 \pmod{p}$ when defined this way. Therefore there is at least one pair (m_1, m_2) of distinct integers in $\mathcal{M}$ such that $\overline{m}_1 + \overline{m}_2 \equiv a \pmod{p^\nu}$, and we can assume that $m_1 < m_2$ by relabeling if necessary. Since $p^\nu - (p+3)/2 \geq (p^\nu - 3)/2$, this establishes the lemma when $p \geq 5$.

On the other hand, if $p = 3$ then we must have $\nu \geq 2$, and the lemma can be shown to hold by letting (m_1, m_2) equal $(3^\nu - 2, 3^\nu - 1)$, $(3^\nu - 4, 3^\nu - 1)$, or $(3^\nu - 5, 3^\nu - 2)$, according to whether a is congruent to 0, 1, or 2 $\pmod{3}$. ∎

The following lemma is one of the two main tools used in our recursive construction in the proof of Proposition 7. This lemma allows us to control all but the smallest prime powers that can appear in the denominators of the rational numbers to be constructed.

LEMMA 15. *Let $q \geq 4$ be a prime power and let c/d be a rational number with $P^*(d) \leq q$. There exists a set $\mathcal{U}$ of integers satisfying*:

(i) $\mathcal{U}$ *is contained in $[q^2/5, q^2]$*;
(ii) $|\mathcal{U}| = 2$ *if q is odd, while $|\mathcal{U}| = 0$ or 1 if q is even*;
(iii) *for each element n of $\mathcal{U}$, $P^*(n) = q$*;
(iv) *if $c'/d' = c/d - \sum_{n \in \mathcal{U}} 1/n$ in lowest terms, then $P^*(d') < q$.*

P r o o f. First assume that q is odd. We apply Lemma 14 with $p^\nu = q$ and

$$a = \begin{cases} c\overline{(d/q)} \pmod{p} & \text{if } q \text{ divides } d, \\ 0 \pmod{p} & \text{if } q \text{ does not divide } d, \end{cases}$$

finding two distinct integers m_1 and m_2 in the range $[(q-3)/2, q)$ such that $\overline{m}_1 + \overline{m}_2 \equiv a \pmod{p}$. Let $\mathcal{U} = \{qm_1, qm_2\}$. Then properties (i)–(iii) are easily seen to hold (the first because $(q-3)/2 \geq q/5$ for $q \geq 5$), and property (iv) holds because of the congruence $\pmod{p}$ satisfied by m_1 and m_2.

On the other hand, if q is even then $q = 2^\nu$ for some $\nu \geq 2$. If 2^ν does not divide d then $P^*(d) < q$ already, and we simply put $\mathcal{U} = \emptyset$. If 2^ν does divide d, then we easily check that the set $\mathcal{U} = \{2^\nu(2^\nu - 1)\}$ satisfies properties (i)–(iv). ∎

The following lemma is the second of the two main tools used in our recursive construction in the proof of Proposition 7. This lemma allows us to control the smallest prime powers that can appear in the denominators of the rational numbers to be constructed. Recall that $L(x) = \text{lcm}\{1, 2, \ldots, \lfloor x \rfloor\}$, as was defined during the proof of Lemma 8.

LEMMA 16. *Let c/d be a rational number and define $q = P^*(d)$, where q is a power of the prime p. There exists an integer n satisfying*:

(i) $L(q)/(p-1) \leq n \leq e^{2q}$;

(ii) $P^*(n) = q$;

(iii) *if $c'/d' = c/d - 1/n$ in lowest terms, then $P^*(d') < q$.*

Proof. Define a to be the residue class of $(L(q)/q)c\overline{(d/q)} \pmod{p}$, where $1 \leq a \leq p - 1$, and let $n = L(q)/a$. Then it can be checked that properties (i) and (ii) hold by the definition of $L(x)$ and the subsequent estimate (12), while property (iii) holds by the choice of a. ∎

Proof of Proposition 7. Define $z = \pi^*(y)$. Our strategy is very similar to the strategy of the proof of Proposition 6. We recursively define a sequence $\{a_i/b_i\}$ $(z + 1 \geq i \geq 1)$ of rationals that decrease in size as the index i decreases, such that the largest prime-power divisor of each b_i is less than q_i. The last member of this sequence, a_0/b_0, will be an integer, and we shall show that it must be zero by bounding its absolute value. The first member a_{z+1}/b_{z+1} will be our original a/b, and each a_i/b_i will be obtained from the previous a_{i+1}/b_{i+1} by subtracting two unit fractions (except in a few cases where we subtract only one or none at all), and the collection of the denominators of all these unit fractions will almost be the desired set $\mathcal{S}$. This collection will have slightly fewer than the desired $2z$ elements, but we shall rectify the error with a simple modification of the splitting identity (2), and the resulting collection will be our set $\mathcal{S}$.

Define $a_{z+1}/b_{z+1} = a/b$ and let $y' = \log y$ and $z' = \pi^*(y')$. We recursively define rationals $\{a_z/b_z, \ldots, a_1/b_1\}$ and sets $\{\mathcal{S}_z, \ldots, \mathcal{S}_1\}$ of integers as follows. If $z' < i \leq z$, then we apply Lemma 15 with $q = q_i$ and $c/d = a_{i+1}/b_{i+1}$. The requirement of the lemma that $P^*(b_{i+1}) \leq q_i$ is satisfied for $i = z$ by the hypothesis of the proposition and the definition of z, and it will be satisfied for smaller values of i by the recursive construction (as we shall see in a moment). Let $\mathcal{S}_i$ be the set $\mathcal{U}$ obtained from applying

Lemma 15, and let

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

in lowest terms, so that by the lemma, $P^*(b_i) < q_i$ and thus $P^*(b_i) \leq q_{i-1}$ (justifying the claim of the previous sentence).

If instead $1 \leq i \leq z'$, then we check whether q_i divides b_{i+1}. If not, then we simply set $\mathcal{S}_i = \emptyset$ and $a_i/b_i = a_{i+1}/b_{i+1}$; since $P^*(b_{i+1}) \leq q_i$ by the recursive construction as before and since q_i does not divide b_{i+1}, we see that $P^*(b_i) \leq q_{i-1}$. On the other hand, if q_i does divide b_{i+1}, then we apply Lemma 16 with $c/d = a_{i+1}/b_{i+1}$ and $q = q_i$. Let $\mathcal{S}_i$ be the set $\{n\}$ where n is the integer obtained from applying Lemma 16, and let

$$\frac{a_i}{b_i} = \frac{a_{i+1}}{b_{i+1}} - \frac{1}{n}$$

in lowest terms, so that by the lemma, $P^*(b_i) < q_i$ and thus $P^*(b_i) \leq q_{i-1}$.

Now set $\mathcal{S}' = \bigcup_{i=1}^{z} \mathcal{S}_i$, so that

$$\frac{a}{b} = \frac{a_0}{b_0} + \sum_{n \in \mathcal{S}'} \frac{1}{n}.$$

We claim that in fact $a_0/b_0 = 0$. It is certainly an integer since its denominator b_0 satisfies $P^*(b_0) \leq q_0 = 1$ by construction, and it is less than 1 by the hypothesis that $a/b < 1$. Moreover, each element of each $\mathcal{S}_i$ is at least $q_i^2/5$ if $i > z'$, and at least $L(q_i)/(p_i - 1)$ if $i \leq z'$, since these sets resulted from applying Lemmas 15 and 16, respectively. Therefore, since $|\mathcal{S}_i| \leq 2$ if $i > z'$ and $|\mathcal{S}_i| \leq 1$ if $i \leq z'$,

$$\frac{a_0}{b_0} = \frac{a}{b} - \sum_{n \in \mathcal{S}'} \frac{1}{n} > \frac{1}{\log y} - 2 \sum_{z' < i \leq z} \frac{5}{q_i^2} - \sum_{1 \leq i \leq z'} \frac{p_i - 1}{L(q_i)}$$

$$= \frac{1}{\log y} - \sum_{\log y < q \leq y} \frac{10}{q^2} - \sum_{1 \leq i \leq z'} \frac{p_i - 1}{L(q_i)}.$$

Since $L(q_i)/p_i = L(q_{i-1})$, this last sum is a telescoping sum whose value is $1 - 1/L(q_{z'})$, as is established by Croot [3, Lemma 1]; in particular, it is less than 1. Also, the penultimate sum is $\ll (\log y \log \log y)^{-1}$ by Mertens' formula (13) and partial summation. Therefore

$$\frac{a_0}{b_0} > \frac{1}{\log y} - O\left(\frac{1}{\log y \log \log y}\right) - 1 > -1$$

when y is sufficiently large. Therefore $a_0/b_0 = 0$, and $a/b = \sum_{n \in \mathcal{S}'} 1/n$.

Again by Lemmas 15 and 16, the members of each $\mathcal{S}_i$ do not exceed

$$\begin{cases} q_i^2 & \text{if } z' < i \leq z \\ e^{2q_i} & \text{if } 1 \leq i \leq z' \end{cases} \leq \max\{y^2, e^{2y'}\} = y^2.$$

In addition, if n is a member of $\mathcal{S}_i$ then $P^*(n) = q_i$, and so the $\mathcal{S}_i$ are pairwise disjoint; the cardinality of each $\mathcal{S}_i$ is 2, except when $1 \leq i \leq z'$ or when q_i is a power of 2, when the cardinality of $\mathcal{S}_i$ is 0 or 1. Since there are $\ll \log y$ of these exceptional values of i, we see that $0 \leq 2z - |\mathcal{S}'| \ll \log y$. Let n be the largest element of $\mathcal{S}'$ and $m = 2z - |\mathcal{S}'|$, and define

$$\mathcal{S} = (\mathcal{S}' \setminus \{n\}) \cup \{n + m, n(n+1), (n+1)(n+2), \ldots, (n+m-1)(n+m)\}.$$

Then the cardinality of $\mathcal{S}$ is exactly $2z = 2\pi^*(y)$, and the largest element of $\mathcal{S}$ is

$$(n + m - 1)(n + m) \leq (y^2 + O(\log y))^2 \leq 2y^4$$

when y is sufficiently large. Moreover, since the identity

$$\frac{1}{n} = \frac{1}{n+m} + \frac{1}{n(n+1)} + \frac{1}{(n+1)(n+2)} + \ldots + \frac{1}{(n+m-1)(n+m)}$$

is valid for any positive integers m and n, we see also that $\sum_{n \in \mathcal{S}} 1/n = \sum_{n \in \mathcal{S}'} 1/n = a/b$. Therefore $\mathcal{S}$ has all of the properties required by Proposition 7.

6. The finiteness of $\mathcal{L}_j(r)$ for $j \geq 2$. In this section we establish Theorem 2. For the reader's convenience we recall the definition of the sets $\mathcal{L}_j(r)$ under consideration:

$$\mathcal{L}_j(r) = \{x \in \mathbb{Z}, \, x > r^{-1} : \text{there do not exist } x_1, \ldots, x_t \in \mathbb{Z},$$
$$x_1 > \ldots > x_t \geq 1 \text{ with } \sum_{i=1}^{t} 1/x_i = r \text{ and } x_j = x\},$$

so that $\mathcal{L}_j(r)$ is the set of numbers that *cannot* be the jth-largest denominator in an Egyptian fraction representation of r. We will make use of the following two lemmas, the first of which is a simple consequence of Proposition 4 stated in a more convenient form.

LEMMA 17. *Let I be a closed subinterval of $(0, \infty)$. There exists a positive real number $X(I)$ such that, for all real numbers $x > X(I)$ and all rational numbers $r = a/b \in I$ for which $P^*(b) < x \log^{-23} x$, there is a set $\mathcal{E}$ of positive integers not exceeding x such that $\sum_{n \in \mathcal{E}} 1/n = r$.*

P r o o f. This follows immediately from Proposition 4 if we set

$$t = \left\lceil (1 - e^{-r})x - \frac{C(I)x \log \log x}{\log x} \right\rceil,$$

where $C(I)$ is a constant that is chosen so large that the right-hand side of the inequality (6) is less than x, and note that both $t > T(I)$ and $t \log^{-22} t > x \log^{-23} x$ will be true as long as $X(I)$ is large enough. ∎

LEMMA 18. *There exists a positive constant k_0 such that, for any integer $k > k_0$, there exists a positive integer $K \equiv -1 \pmod{k}$ such that $P^*(K) < k \log^{-24} k$.*

Because of its length and technical nature, we defer the proof of Lemma 18 until the end of this section. Assuming this lemma to be true, we may now proceed with a proof of Theorem 2.

Proof of Theorem 2. We begin by showing that $\mathcal{L}_2(r)$ is finite for any r. Let $r = a/b$ be a positive rational number, and set $I = [r/2, r]$. Let k be any integer satisfying $k \log^{-24} k > P^*(b)$ and $k > \max\{4/r, X(I), k_0\}$, where $X(I)$ and k_0 are the constants described in Lemmas 17 and 18, respectively. We claim that there exists an Egyptian fraction representation of r whose second-largest denominator is k, and hence that $k \notin \mathcal{L}_2(r)$.

To see this, let K be a positive integer such that $K \equiv -1 \pmod{k}$ and $P^*(K) < k \log^{-24} k$, as guaranteed by Lemma 18. Let $x = k/\log k$ and define the rational number $r' = a'/b'$ by

$$r' = r - \frac{1}{k} - \frac{1}{Kk}.$$

We have that $1/k + 1/(Kk) < 2/k < r/2$, and hence r' is in I. Also,

$$\frac{1}{k} + \frac{1}{Kk} = \frac{K+1}{Kk} = \frac{(K+1)/k}{K},$$

where the numerator is an integer since $K \equiv -1 \pmod{k}$, and so

$$P^*(b') \leq \max\{P^*(b), P^*(K)\} \leq k \log^{-24} k < x \log^{-23} x.$$

Therefore we can invoke Lemma 17 with r' and x to produce a set $\mathcal{E}$ of positive integers not exceeding x such that $\sum_{n \in \mathcal{E}} 1/n = r'$, whence $\mathcal{E} \cup \{k, Kk\}$ is the set of denominators for an Egyptian fraction representation of r with second-largest denominator equal to k. Therefore k is not an element of $\mathcal{L}_2(r)$, and since this argument holds for all k that are sufficiently large in terms of r as specified above (note that the constant $X(I)$ depends only on r), we have shown that $\mathcal{L}_2(r)$ is finite.

Now that we know that $\mathcal{L}_2(r)$ is finite, we can establish Theorem 2 in its full strength. First, notice that if there exists an Egyptian fraction representation of r with the integer n as its jth-largest denominator ($j \geq 2$), then by splitting the term with largest denominator using the identity (2), we easily obtain an Egyptian fraction representation of r with n as its $(j+1)$st-largest denominator. This shows that $\mathcal{L}_2(r) \supset \mathcal{L}_3(r) \supset \ldots$ In particular, since $\mathcal{L}_2(r)$ is finite, it follows that all of the $\mathcal{L}_j(r)$ are finite for $j \geq 3$.

Furthermore, for every element n of $\mathcal{L}_2(r)$, we can find an Egyptian fraction representation of $r - 1/n$ using only denominators exceeding n (there are many ways to do this—one could use a greedy algorithm, for example). If $r - 1/n = 1/n_1 + \ldots + 1/n_{j-1}$ is such a representation, then the representation $r = 1/n_1 + \ldots + 1/n_{j-1} + 1/n$ of r shows that $n \notin \mathcal{L}_j(r)$. Since we can find such an integer j for each $n \in \mathcal{L}_2(r)$, and since the $\mathcal{L}_j(r)$ form a nested decreasing sequence of sets each contained in $\mathcal{L}_2(r)$, we see that at some

point the sets $\mathcal{L}_j(r)$ will be empty. This completes the proof of Theorem 2 (modulo the proof of Lemma 18). ■

We remark that our proof does not show that $\mathcal{L}_2(r) \subset \mathcal{L}_1(r)$, and indeed this is false in general—in fact it is not even the case that $\mathcal{L}_3(r) \subset \mathcal{L}_1(r)$ always. For example, it is easy to see that if p is a prime, then $p+1$ cannot be the third-largest denominator in an Egyptian fraction representation of $r = 1/p + 1/(p+1)$, and so $p+1$ is an element of $\mathcal{L}_3(r)$ (hence of $\mathcal{L}_2(r)$ as well) but not $\mathcal{L}_1(r)$. (On the other hand, it seems likely that $\mathcal{L}_2(1) \subset \mathcal{L}_1(1)$, for instance, although this does not seem trivial to show.) We can always convert an Egyptian fraction representation of r whose largest denominator is some integer n into one whose fourth-largest denominator is n, by repeatedly splitting the term with largest denominator other than n and examining the various ways in which the term $1/n$ could be duplicated under this process. In this way one can show that $\mathcal{L}_4(r) \subset \mathcal{L}_1(r)$ for every r. It might be interesting to try to classify the rational numbers r for which $\mathcal{L}_2(r) \not\subset \mathcal{L}_1(r)$.

We now return to the task of establishing Lemma 18. One possibility would be to cite an existing result on smooth numbers in arithmetic progressions in which the modulus of the progression was allowed to exceed the smoothness parameter, such as a theorem of Balog and Pomerance [1], and remove those numbers divisible by large prime powers in an *ad hoc* manner. We prefer to provide a self-contained proof of Lemma 18, one that nevertheless has ideas in common with the method of Balog and Pomerance, including a reliance on estimates for incomplete Kloosterman sums.

The following lemma is still much stronger than we need but nearly the least we could prove to establish Lemma 18. The author would like to thank Henryk Iwaniec for a helpful conversation concerning the proof of this lemma.

LEMMA 19. *Let k be a positive integer and ε and x positive real numbers such that $k^{6/7+\varepsilon} < x \le k$. The number of ordered pairs (m, n) of coprime positive integers less than x such that $mn \equiv -1 \pmod{k}$ is*

$$(26) \qquad \frac{6x^2}{\pi^2 k} \prod_{p \mid k} \left(\frac{p}{p+1} \right) + O_\varepsilon(x^{5/6} k^\varepsilon).$$

In particular, such ordered pairs (m, n) exist when k is sufficiently large in terms of ε.

Of course, Lemma 18 follows by setting $\varepsilon = 1/14$, say, choosing x such that $k^{13/14} < x < k \log^{-24} k$, and letting $K = mn$ where (m, n) is one of the pairs whose existence is ensured by Lemma 19 when k is sufficiently large. The restriction that m and n be coprime causes a little more trouble than one might think; without it, one could simplify the proof and easily obtain

an error term of $xk^{-1/4+\varepsilon}$ in the asymptotic formula (26), whence x could be taken as small as $k^{3/4+\varepsilon}$ in the hypotheses of the lemma. In any case, even a hypothesis as weak as $x > k^{1-\delta}$ would be ample for our needs, so we have not gone to great lengths to make the error terms as small as possible in the proof.

Proof of Lemma 19. All of the constants implicit in the $\ll$- and O-notation in this proof may depend on ε. We use $d(n)$ to denote the number of divisors of n and note that $d(n) \ll n^\varepsilon$ for any positive ε. We also recall that $\|x\|$ denotes the distance from x to the nearest integer and that $e_k(x) = e^{2\pi ix/k}$.

Let N denote the number of ordered pairs (m, n) of coprime positive integers less than x such that $mn \equiv -1 \pmod k$, so that

$$N = \sum_{\substack{m \le x \\ (m,n)=1 \\ mn \equiv -1 \,(\mathrm{mod}\,k)}} \sum_{n \le x} 1 = \sum_{\substack{m \le x \\ mn \equiv -1 \,(\mathrm{mod}\,k)}} \sum_{n \le x} \sum_{d|(m,n)} \mu(d) = \sum_{d \le x} \mu(d) \sum_{\substack{m \le x/d \\ mnd^2 \equiv -1 \,(\mathrm{mod}\,k)}} \sum_{n \le x/d} 1$$

by changing m and n to md and nd, respectively. Let y and z be real numbers to be specified later subject to $1 \le z \le y \le x$, and write $N = N_1 + O(N_2 + N_3)$ where

$$N_1 = \sum_{d \le z} \mu(d) \sum_{\substack{m \le x/d \\ mnd^2 \equiv -1 \,(\mathrm{mod}\,k)}} \sum_{n \le x/d} 1,$$

$$N_2 = \sum_{z < d \le y} \sum_{\substack{m \le x/d \\ mnd^2 \equiv -1 \,(\mathrm{mod}\,k)}} \sum_{n \le x/d} 1,$$

$$N_3 = \sum_{y < d \le x} \sum_{\substack{m \le x/d \\ mnd^2 \equiv -1 \,(\mathrm{mod}\,k)}} \sum_{n \le x/d} 1 \le \sum_{m \le x/y} \sum_{n \le x/y} \sum_{\substack{y < d \le x \\ mnd^2 \equiv -1 \,(\mathrm{mod}\,k)}} 1.$$

We begin by bounding N_2 and N_3. The estimation of N_2 is trivial: writing $l = mn$ we have

$$N_2 \le \sum_{z < d \le y} \sum_{\substack{l \le x^2/d^2 \\ ld^2 \equiv -1 \,(\mathrm{mod}\,k)}} d(l) \ll (x^2)^{\varepsilon/2} \sum_{z < d \le y} \left(\frac{x^2}{d^2 k} + 1\right) \ll \left(\frac{x^2}{zk} + y\right) k^\varepsilon.$$

As for N_3, we use Cauchy's inequality to write

$$N_3^2 \le \frac{x^2}{y^2} \sum_{m \le x/y} \sum_{n \le x/y} \left(\sum_{\substack{y < d \le x \\ mnd^2 \equiv -1 \,(\mathrm{mod}\,k)}} 1 \right)^2 = \frac{x^2}{y^2} \sum_{m \le x/y} \sum_{n \le x/y} \sum_{\substack{y < d_1 \le x \\ mnd_1^2 \equiv -1 \,(\mathrm{mod}\,k)}} \sum_{\substack{y < d_2 \le x \\ mnd_2^2 \equiv -1 \,(\mathrm{mod}\,k)}} 1.$$

The congruence conditions imply that m and n must be coprime to k and thus that $d_1^2 \equiv d_2^2 \pmod{k}$. Therefore we can weaken the conditions on the variables d_1 and d_2 and make them independent of m and n, yielding the upper bound

$$(27) \qquad N_3^2 \leq \frac{x^2}{y^2} \sum_{m \leq x/y} \sum_{n \leq x/y} \sum_{\substack{y < d_1 \leq x}} \sum_{\substack{y < d_2 \leq x \\ d_1^2 \equiv d_2^2 \,(\mathrm{mod}\,k)}} 1$$

$$\leq \frac{x^4}{y^4} \sum_{\substack{|l| \leq x^2 \\ k|l}} \#\{1 \leq d_1, d_2 \leq x : d_1^2 - d_2^2 = l\}$$

by writing $l = d_1^2 - d_2^2$. The term $l = 0$ contributes $\lfloor x \rfloor$, while the remaining terms contribute at most $d(|l|)$ each, since d_1 and d_2 are determined by $d_1 + d_2$ and $d_1 - d_2$ which must be complementary divisors of l. Because $|l| \leq x^2 \leq k^2$, we see that $d(|l|) \ll k^\varepsilon$. Since the number of terms with $|l| \neq 0$ in the final sum in equation (27) is $\ll x^2/k \leq x$, the sum is bounded by $O(xk^\varepsilon)$, and so equation (27) gives us

$$N_3 \ll \left(\frac{x^4}{y^4} x k^\varepsilon \right)^{1/2} \leq \frac{x^{5/2} k^\varepsilon}{y^2}.$$

We have thus shown that

$$(28) \qquad N = N_1 + O\left(\left(\frac{x^2}{zk} + y + \frac{x^{5/2}}{y^2} \right) k^\varepsilon \right).$$

It remains to evaluate N_1. Using the additive characters $e_k(x)$ to detect the condition $mnd^2 \equiv -1 \pmod{k}$, or equivalently $(dm, k) = 1$ and $\overline{m} + nd^2 \equiv 0 \pmod{k}$, we have

$$(29) \quad N_1 = \sum_{\substack{d \leq z \\ (d,k)=1}} \mu(d) \sum_{\substack{m \leq x/d \\ (m,k)=1}} \sum_{n \leq x/d} \frac{1}{k} \sum_{h\,(\mathrm{mod}\,k)} e_k(h(\overline{m} + nd^2))$$

$$= \frac{1}{k} \sum_{\substack{d \leq z \\ (d,k)=1}} \mu(d) \sum_{h\,(\mathrm{mod}\,k)} \left(\sum_{\substack{m \leq x/d \\ (m,k)=1}} e_k(h\overline{m}) \right) \left(\sum_{n \leq x/d} e_k(hnd^2) \right).$$

The terms with $h \equiv 0 \pmod{k}$ contribute

$$\frac{1}{k} \sum_{\substack{d \leq z \\ (d,k)=1}} \mu(d) \left(\sum_{\substack{m \leq x/d \\ (m,k)=1}} 1 \right) \left(\sum_{n \leq x/d} 1 \right)$$

$$= \frac{1}{k} \sum_{\substack{d \leq z \\ (d,k)=1}} \mu(d) \left(\frac{x\phi(k)}{dk} + O(k^{\varepsilon/2}) \right) \left(\frac{x}{d} + O(1) \right)$$

$$= \frac{x^2 \phi(k)}{k^2} \sum_{\substack{d \leq z \\ (d,k)=1}} \frac{\mu(d)}{d^2} + O\left(\frac{x}{k^{1-\varepsilon/2}} \sum_{\substack{d \leq z \\ (d,k)=1}} \frac{1}{d}\right)$$

$$= \frac{x^2 \phi(k)}{k^2} \left(\frac{1}{\zeta(2)} \prod_{p|k}\left(1 - \frac{1}{p^2}\right)^{-1} + O\left(\frac{1}{z}\right)\right) + O\left(\frac{x}{k^{1-\varepsilon}}\right).$$

Both error terms are $O(x^2/(zk^{1-\varepsilon}))$, and so equation (29) becomes

$$(30) \qquad N_1 = \frac{x^2 \phi(k)}{\zeta(2)k^2} \prod_{p|k}\left(1 - \frac{1}{p^2}\right)^{-1} + O\left(\frac{x^2}{zk^{1-\varepsilon}} + \frac{1}{k}\sum_{\substack{d \leq z \\ (d,k)=1}} T(d)\right),$$

where we have defined

$$T(d) = \sum_{h \not\equiv 0 \,(\mathrm{mod}\, k)} \left(\sum_{\substack{m \leq x/d \\ (m,k)=1}} e_k(h\overline{m})\right)\left(\sum_{n \leq x/d} e_k(hnd^2)\right).$$

To estimate the $T(d)$, we make use of the elementary bound

$$\sum_{n \leq x/d} e_k(hnd^2) \ll \left\|\frac{hd^2}{k}\right\|^{-1}$$

when k does not divide hd^2, and also the Weil bound for incomplete Kloosterman sums (see for instance Hooley [7, Lemma 4 of Section 2.5]), which gives as a special case

$$\sum_{\substack{m \leq x/d \\ (m,k)=1}} e_k(h\overline{m}) \ll_\varepsilon k^{1/2+\varepsilon/2}(h,k)^{1/2}.$$

It follows that

$$(31) \qquad T(d) \ll k^{1/2+\varepsilon/2} \sum_{h \not\equiv 0 \,(\mathrm{mod}\, k)} (h,k)^{1/2}\left\|\frac{hd^2}{k}\right\|^{-1}$$

when d is coprime to k. Because $(d,k) = 1$ we may reindex this sum by replacing hd^2 with h, which does not affect the quantity (h,k). When $1 \leq h \leq k/2$, we note that $\|h/k\|^{-1} = \|(k-h)/k\|^{-1} = k/h$ and $(h,k) = (k-h,k)$, and so

$$\sum_{h \not\equiv 0 \,(\mathrm{mod}\, k)} (h,k)^{1/2}\left\|\frac{hd^2}{k}\right\|^{-1} = 2k \sum_{1 \leq h \leq k/2} \frac{(h,k)^{1/2}}{h}.$$

By writing $h = h'f$ where $f = (h,k)$, this last sum is easily seen to be $\ll d(k)\log k$, and $T(d) \ll k^{3/2+\varepsilon}$ follows from this bound and the estimate (31).

Using this estimate for the $T(d)$ in equation (30), the asymptotic formula (28) for N becomes

$$(32) \quad N = \frac{x^2 \phi(k)}{\zeta(2)k^2} \prod_{p|k} \left(1 - \frac{1}{p^2}\right)^{-1} + O\left(\left(\frac{x^2}{zk} + zk^{1/2} + y + \frac{x^{5/2}}{y^2}\right)k^\varepsilon\right).$$

We optimize this error term by choosing $z = xk^{-3/4}$ and $y = x^{5/6}$, in which case the error term is $O((xk^{-1/4} + x^{5/6})k^\varepsilon)$, and the $x^{5/6}k^\varepsilon$ term dominates since $x \le k$. We also note that $\zeta(2) = \pi^2/6$ and that

$$\frac{\phi(k)}{k} \prod_{p|k} \left(1 - \frac{1}{p^2}\right)^{-1} = \prod_{p|k} \left(\frac{p}{p+1}\right),$$

and so the main term in (32) is the same as the main term in (26). This establishes the lemma.

7. The order of magnitude of $L_1(r; x)$. In this section we establish Theorem 3. We recall that $L_1(r; x)$ is the number of integers in $\mathcal{L}_1(r)$ not exceeding x, that is, the number of integers not exceeding x that *cannot* be the largest denominator in an Egyptian fraction representation of r.

Proof of Theorem 3. First we establish the lower bound in the inequality (4). Let r be a positive rational number and $x > 1$ a real number, set $y = Cx/\log x$ with $C > 1$ a large constant, and suppose that x is so large that all prime divisors of the denominator of r are less than y. By Lemma 8, if $n \le x$ is the largest denominator in an Egyptian fraction representation of r, it must be true that n is not divisible by any prime larger than y (when C is chosen large enough). In other words, the set $\mathcal{L}_1(r)$ contains all integers $n \le x$ such that $P(n) > y$. The number of integers $n \le x$ such that $P(n) > y$ is asymptotic to $x \log\log x/\log x$ by Lemma 9, which establishes the required lower bound.

To establish the corresponding upper bound, we use a method similar to the one used in the proof of Theorem 2 to argue that $\mathcal{L}_2(r)$ is finite. Let $r = a/b$ be a positive rational number, and set $I = [r/2, r]$. Let x be a positive real number and set $x' = x/\log x$ and $y = x' \log^{-23} x'$, so that $y > x \log^{-24} x$. We suppose that x is so large that $y > P^*(b)$ and $x' > \max\{2/r, X(I)\}$, where $X(I)$ is the constant described in Lemma 17. Let k be an integer such that $x' < k \le x$ and $P^*(k) < y$, and define the rational number $r' = a'/b'$ by $r' = r - 1/k$.

Since $1/k < 1/x' < r/2$ we have $r' \in I$; next, $P^*(b') \le \max\{P^*(b), P^*(k)\} < y$. Therefore we may invoke Lemma 17 with r' and x' to produce a set $\mathcal{E}$ of positive integers not exceeding x' such that $\sum_{n \in \mathcal{E}} 1/n = r'$, whence $\mathcal{E} \cup \{k\}$ is the set of denominators for an Egyptian fraction representation of r with largest denominator k.

From this argument we deduce that for x sufficiently large, the elements of $\mathcal{L}_1(r)$ not exceeding x are all contained in $\{r^{-1} < n \leq x'\} \cup \{r^{-1} < n \leq x : P^*(n) > x \log^{-24} x\}$, whose cardinality is

$$\ll x' + x \log \frac{\log x}{\log y} \ll \frac{x \log \log x}{\log x}$$

by Lemma 9. This upper bound holds when x is large enough in terms of r as described above, but it will hold for all x by adjusting the implicit constant (depending on r) if necessary. This establishes the upper bound in the inequality (4) and hence Theorem 3. ∎

We remark that we have actually established the inequalities

$$\frac{(1 + o_r(1))x \log \log x}{\log x} \leq L_1(r; x) \leq \frac{(24 + o_r(1))x \log \log x}{\log x}.$$

We did so by showing essentially that an integer n is in $\mathcal{L}_1(r)$ if and only if n can be written as $n = p^\nu m$ with m less than a certain power of $\log p$, though we were unable to pinpoint this power other than to show that it lies between $1 + o(1)$ and $24 + o(1)$. With much more care we could improve the constant 24 to 3 but no further at present. Nevertheless we speculate that the correct power is 1, i.e., that for any fixed r and $\varepsilon > 0$ there are only finitely many integers $n \in \mathcal{L}_1(r)$ such that, if $P^*(n) = p^\nu$ and $n = p^\nu m$, then $m \geq \log^{1+\varepsilon} p$. This would imply that the counting function $L_1(r; x)$ of those integers that cannot be the largest denominator in an Egyptian fraction representation of r is asymptotic to $x \log \log x / \log x$ with leading coefficient 1.

References

[1] A. Balog and C. Pomerance, *The distribution of smooth numbers in arithmetic progressions*, Proc. Amer. Math. Soc. 115 (1992), 33–43.

[2] E. S. Croot III, *On unit fractions with denominators from short intervals*, Acta Arith., to appear.

[3] —, *On some questions of Erdős and Graham about Egyptian fractions*, Mathematika, to appear.

[4] P. Erdős and R. L. Graham, *Old and New Problems and Results in Combinatorial Number Theory*, Monograph. Enseign. Math. 28, Univ. Genève, Geneva, 1980.

[5] R. K. Guy, *Unsolved Problems in Number Theory*, 2nd ed., Springer, New York, 1994.

[6] A. Hildebrand and G. Tenenbaum, *Integers without large prime factors*, J. Théor. Nombres Bordeaux 5 (1993), 411–484.

[7] C. Hooley, *Applications of Sieve Methods to the Theory of Numbers*, Cambridge Tracts in Math. 70, Cambridge Univ. Press, Cambridge, 1976.

[8] G. Martin, *Dense Egyptian fractions*, Trans. Amer. Math. Soc. 351 (1999), 3641–3657.

[9]   R. C. V a u g h a n, *The Hardy–Littlewood Method*, 2nd ed., Cambridge Univ. Press, Cambridge, 1997.
[10] H. Y o k o t a, *On a problem of Erdős and Graham*, J. Number Theory 39 (1991), 327–338.

Department of Mathematics
University of Toronto
Canada M5S 3G3
E-mail: gerg@math.toronto.edu