ERNEST S. CROOT III
Unit Fractions
(Under the direction of ANDREW GRANVILLE)

We will give some of the history of the theory of Unit Fractions, and will state and prove the following three results, which answer previously unsolved problems of Paul Erdős and R. L. Graham:

1. For N sufficiently large, every integer m with

$$1 \leq m \leq \left[\sum_{1 \leq n \leq N} \frac{1}{n} - \left(\frac{9}{2} + o(1) \right) \frac{(\log \log N)^2}{\log N} \right]$$

can be written as $m = \sum_{1 \leq n \leq N} \epsilon_n / n$, where $\epsilon_n = 0$ or 1.

2. For any rational $r > 0$ and all $N > 1$, there exist integers $x_1, x_2, ..., x_k$, where

$$N < x_1 < x_2 < \cdots < x_k \leq \left(e^r + O_r \left(\frac{\log \log N}{\log N} \right) \right) N$$

such that

$$r = \frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_k}.$$

3. There exists a constant $b > 0$ such that if we r-color the integers in $[2, b^r]$, then there exists a monochromatic set S such that $\sum_{n \in S} 1/n = 1$.

INDEX WORDS: Unit Fractions, Egyptian Fractions, Coloring Conjecture, Erdős

Unit Fractions

by

Ernest S. Croot III

B.S., Centre College, 1994

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2000

Unit Fractions

by

Ernest S. Croot III

Approved:

Major Professor

Date

Approved:

Graduate Dean

Date

Acknowledgements

I would like to thank Andrew Granville, my advisor, for encouraging me to obtain stronger and stronger versions of the results in this thesis, and for helping me to edit the papers which comprise Chapters 4,5, and 6, to get them into their current form; Carl Pomerance and Greg Martin, for their enlightening conversations and suggestions; Christian Elsholtz, for his many suggestions and corrections to the first two papers in this work; and finally, I would like to thank R. L. Graham and Paul Erdős, for the wonderful questions which inspired most of the work in this dissertation.

Contents

Chapter 1

Introduction

While on a trip to Egypt, in 1858 the Scottish Egyptologist Alexander Henry Rhind purchased several ancient Egyptian artifacts at a market in Luxor, including a papyrus roll, which we now call the Rhind Mathematical Papyrus (see [26]). The writing on the papyrus was in a hieratic script, a cursive form of hieroglyphics, and began with the following message:[1]

> *Correct method of reckoning, for grasping the meaning of things and knowing everything that is, obscurities and all secrets.*

Below this was written the date, which revealed that the papyrus was written around 1650 B.C., and the name Ahmose, who claimed to be the scribe that had copied its text from a much earlier work. [2]

The next section of the papyrus had instructions for how to manipulate unit fractions, which are numbers of the form $1/n$, where n is a positive integer, and contained many puzzles of a practical nature illustrating their use. Among these instructions was a lookup table for how to express numbers of the form $2/n$, where $3 \leq n \leq 101$ and n is odd, as a sum of at most four distinct unit fractions. [3] This

[1]The papyrus was divided into two parts: the Recto, and the Verso. The introductory matter was written on the Recto, which, along with the the Verso, contained several mathematical problems.

[2]Ahmose claims this earlier work comes from the time of King Ny-maat-re, who ruled duing the second half of the 19th century B.C.

[3]Curiously, the Egyptians did not always bother with decomposing $2/3$ as a sum of unit fractions, and they even had a special symbol to denote it.

table, in turn, was used to represent rational numbers as a sum of unit fractions, which was a fundamental operation in Egyptian arithmetic.

An example of how the Egyptians used unit fractions for practical calculations (more precisely, how they used unit fractions to represent the answer to a practical problem), as well as the table for expanding $2/n$, is illustrated by problem number 65 of the Papyrus, which is as follows (very loosely translated):

> 100 loaves of bread are to be divided among ten men, three of whom are a sailor, a foreman, and a watchman, and who get double the share of the other seven. How many loaves do the seven get, and how many do the three get?

The number of portions of bread is $13 = 10$ single portions $+\ 3$ extra portions, and so the seven men would each get $100/13 = 7 + 9/13$ loaves. The ancient Egyptians would not be satisfied with the remainder expressed as $9/13$; they would try to further decompose this into a sum of unit fractions, and had many ways of doing so. A common method used was as follows: To expand $A/B < 1$ as a sum of unit fractions, where $\gcd(A, B) = 1$, first, find a set of integers $1 \leq a_1 < a_2 < \cdots < a_k$, where

$$A - 2 < B \sum_{j=1}^{k} \frac{1}{2^{a_j}} \leq A.$$

Then, we have

$$A - B \sum_{j=1}^{k} \frac{1}{2^{a_j}} = \sum_{j=1}^{l} \frac{1}{2^{b_j}},$$

for some $0 \leq b_1 < \cdots < b_l \leq a_k$, so that

$$\frac{A}{B} = \sum_{j=1}^{k} \frac{1}{2^{a_j}} + \sum_{j=1}^{l} \frac{1}{B 2^{b_j}}.$$

A variation on this method, which was also used by the Egyptians, was to find $0 \le a'_1 < \cdots < a'_m$ and $1 \le b'_1 < \cdots < b'_n \le a'_m$, where

$$A - \frac{2}{3}B\sum_{j=1}^{m}\frac{1}{2^{a'_j}} = \frac{2}{3}\sum_{j=1}^{n}\frac{1}{2^{b'_j}},$$

so that

$$A = \frac{2}{3}\left\{\sum_{j=1}^{m}\frac{1}{2^{a'_j}} + \sum_{j=1}^{n}\frac{1}{B2^{b'_j}}\right\}.$$

(Note that this second expansion for A/B is not necessarily a sum of unit fractions, because it can contain the terms 2/3. The ancient Egyptians treated 2/3 as though it were a unit fraction, and so they did not mind having it in expansions.) In the case of our bread division problem where $A/B = 9/13$, this second method was used, and the expansion obtained was $9/13 = 2/3 + 1/39$. Thus, the seven men who got a single portion of bread each received $7 + 2/3 + 1/39$ loaves, and the other three men received double this, which is $15 + 1/3 + 2/39$. The unit fraction expansion for 2/39 was then looked up in the table, and be found to equal $1/26 + 1/78$. Thus, the three men each received $15 + 1/3 + 1/26 + 1/78$ loaves.

It seems clear that for much larger problems, unit fractions are not a very efficient way of solving practical problems, at least not in the way that the ancient Egyptians used them. Despite this, they were used for practical calculations well past the time of Ahmose, as is evident in the Akhmim Papyrus, which was written in the sixth century A.D. (see [3]). Today, however, they are mainly a source of mathematical curiosity, and have given rise to many conjectures, most of which are still unanswered, as we will see later on.

Even though the Egyptians worked with unit fractions for millennia, Leonardo Fibonacci is credited (see [12] and [10]) with "proving", in 1202, the first fundamental result in the subject: [4] Every positive rational number ≤ 1 can be written as a sum

[4]Fibonacci's actual name was Leonardo Pisano. Although Fibonacci did not have access to the Rhind Papyrus, unit fractions were known and used during his time; also, he was an

of distinct unit fractions. The more general case, where the rational is allowed to be greater than 1, can be proved using his method, together with the fact that the harmonic series diverges. The main observation Fibonacci made was as follows: Given a rational number $A/B \leq 1$ (which is not already a unit fraction), $\gcd(A, B) = 1$, write it as $A/(An+r)$, where $1 \leq r \leq (A-1)$. Now, $1/(n+1)$ is the largest unit fraction which leaves a non-negative remainder when subtracted from A/B, and if we perform this subtraction, we obtain

$$\frac{A}{An+r} - \frac{1}{n+1} = \frac{A-r}{(An+r)(n+1)}.$$

The numerator of this difference is less than A, and so if one repeats this process of subtracting off the largest unit fraction which leaves a non-negative remainder, one must eventually end up with a remainder which is itself a unit fraction, because the numerators of the successive remainders decrease in size. All the unit fractions which were subtracted will be distinct, and their sum equal to A/B.

From Fibonacci's method, one can show that every positive rational of the form $k/B \leq 1$ can be expressed as a sum of at most k distinct unit fractions. This leads one to wonder whether, for instance, if $(k+1)/B$ can always be written as the sum of at most k distinct unit fractions, when B is sufficiently large. This is clearly not possible when $k = 1$, since $2/B$ is not a unit fraction for any B odd, and for $k = 2$, we have that $3/p$, where $p = 6n+1$ is prime, cannot be expressed as a sum of 2 unit fractions. To see this, suppose that $3/p = 1/x_1 + 1/x_2$, where $x_1 < x_2$. Then, $p \nmid x_1$, else $1/x_1 + 1/x_2 < 2/p < 3/p$. This gives us that $1/x_2 = 3/p - 1/x_1 = (3x_1 - p)/(px_1)$ (note that this is in lowest terms), which implies that $3x_1 - p = 1$ and $px_1 = x_2$; however, this is impossible because $3x_1 - p \equiv -1 \pmod 3$. The case $k = 3$ is the first non-trivial case, and is virtually the same as an old question of Erdős and Straus (see [15] and [12]), who asked whether $4/B = 1/x_1 + 1/x_2 + 1/x_3$ has a solution in the

avid scholar of Greek mathematics, through which a lot of Egyptian mathematical ideas survived.

positive integers for every $B \geq 2$, where here we do not care whether x_1, x_2, and x_3 are distinct. This question remains a major unsolved problem in the subject, though there is some computational and theoretical evidence to support it. For example, the conjecture has been verified for all $B < 10^{14}$ by A. Swett,[5] and holds for every $B \geq 2$, except possibly when $B \equiv 1, 11^2, 13^2, 17^2, 19^2$ or $23^2 \pmod{840}$ (see [23] and [15]). Schinzel and Sierpiński (see [29]) made the more general conjecture that $A/B = 1/x_1 + 1/x_2 + 1/x_3$ is always solvable for $B > B_0(A)$. In [30], Vaughan proved that the number of rationals A/B which do not have such a solution (x_1, x_2, x_3), where A is fixed and B ranges over the integers $\leq N$, is $\ll N \exp(-c(\log N)^{2/3})$, where c is a constant depending only on A. Viola, Shen, and Elsholtz (see [31], [28], [11]) have generalized Vaughan's result, with Elsholtz's work being the most recent. He considers solutions to $A/B = 1/x_1 + \cdots + 1/x_k$, and shows that for a fixed A, the number of exceptional $B \leq N$ is $\ll N/\exp(c_{A,k}(\log N)^{h(k)})$, where $k \geq 3$, $h(k) = 1 - 1/(2^{k-1} - 1)$, and $c_{A,k}$ is a positive constant depending only on A and k. In the case $k = 2$ Hofmeister and Stoll (see [17]) showed that the number of exceptional $B \leq N$ is $\ll N(\log N)^{-1/\phi(A)}$.

In a paper by myself, Dobbs, Hetzel, Friedlander, and Pappalardi (see [6]) we consider the question of estimating the number of integers m coprime to n, such that $m/n = 1/x_1 + \cdots + 1/x_k$. We conjecture that this number is $n^{o(1)}$, but we were only able to show that it is $\ll n^{\alpha_k + \epsilon}$, for every $\epsilon > 0$, where α_k is recursively defined by $\alpha_2 = 0$ and $\alpha_k = 1 - (1 - \alpha_{k-1})/(2 + \alpha_{k-1})$. For the case $k = 2$ we show that, on average, the number of such m is $\asymp \log^3 n$.

If one modifies Fibonacci's method a bit, so that for $A/B = 1$ we start with the fraction $1/2$, and then successively add to this the next smallest unit fraction which

[5] Elsholtz personally communicated this result to me.

gives the closest underapproximation to 1, then one arrives at the expansion:[6]

$$1 = \sum_{j=1}^{\infty} \frac{1}{u_{j+1}}, \text{ where } u_1 = 1,\ u_{j+1} = u_j^2 - u_j + 1 = u_1 u_2 \cdots u_j + 1.$$

J. J. Sylvester investigated this series,[7] and noticed that one can produce, for any $k \geq 3$, the related, finite expansion:

$$1 = \sum_{j=1}^{k-1} \frac{1}{u_{j+1}} + \frac{1}{u_k - 1}.$$

(Note: if $k = 2$, this would give the expansion $1 = 1/2 + 1/2$, which we do not consider, since it consists of a repeated unit fraction.) One can see from the relation $u_{j+1} = u_j^2 - u_j + 1$ that the sequence of u_j's grows quite fast (double exponentially), and so one might suspect that if 1 were written as the sum of k distinct unit fractions, then the largest denominator is always $\leq u_k - 1$. O. D. Kellogg (see [18]) claimed to have a proof that this is the case, though he did not publish it in its entirety, but in [7], D. R. Curtiss supplied the first complete proof.

Let $R(k)$ denote the number of representations $1 = 1/x_1 + \cdots + 1/x_k$, $x_1 < x_2 < \cdots < x_k$. Curtiss's result implies a large, but finite, upper bound for the size of $R(k)$, and one might wonder exactly how large it can be. Erdős and Straus (see [12]) proved that

$$e^{k^{2-o(1)}} < R(k) \leq a^{2^{k+1}},$$

where $a = \lim_{n\to\infty} u_n^{1/2^n} = 1.264085....$ In 1972, David Singmaster calculated the first 6 values for $R(k)$ (see [15]), which suggest that the Erdős-Straus upper bound for $R(k)$ is closer to the truth than is the lower bound: $R(1) = 1$, $R(2) = 1$, $R(3) = 10$, $R(4) = 215$, $R(5) = 12231$, and $R(6) = 2025462$. Erdős asked for an

[6] Here, we do not stop when the remainder is a unit fraction, as in Fibonacci's method. So, if the first few unit fractions which are subtracted from the successive remainders are $1/u_1, 1/u_2, ..., 1/u_k$, and if $1 - 1/u_1 - 1/u_2 - \cdots - 1/u_k = 1/n$, for some n, then we set $u_{k+1} = n + 1$.

[7] He also rediscovered Fibonacci's work, though one could argue that Fibonacci's "proof" does not meet the standards accepted by mathematicians today.

asymptotic estimate for $R(k)$, which remains an unsolved question. Define $R^*(k)$ to be the number of solutions to $1 = 1/x_1 + \cdots + 1/x_n$, for some $n < k$, where here the x_j's are allowed to be equal. One can easily show the $R^*(k) > e^{k^{2-o(1)}}$ by applying the following result, which is due to Vose (see [32]): Given any rational number $A/B > 0$, there exists an expansion $A/B = 1/x_1 + \cdots + 1/x_n$, with $x_1 < \cdots < x_n$ and $n < c\sqrt{\log B}$, for some $c > 0$. To see how this gives the lower bound for $R^*(k)$, we use it to form unit fraction expansions of each of the numbers $(B-1)/B$ where $2 \le B < e^{(k/c)^2}$, and for each such expansion, we will have

$$1 = \frac{1}{B} + \frac{B-1}{B} = \frac{1}{B} + \sum_{j=1}^{n} \frac{1}{x_j}, \text{ where } n < c\sqrt{\log B} < k.$$

Since each expansion corresponds to at most k different values of B, we have that $R^*(k) > e^{(k/c)^2}/k = e^{k^{2-o(1)}}$. Perhaps this simple proof can be modified to show $R(k) > e^{k^{2-o(1)}}$.

On the opposite extreme from the result of Curtiss, Erdős asked how small the largest denominator can possibly be if 1 is expressed as a sum of k distinct unit fractions. One can show that this largest denominator is no smaller than $e(k + o(1))/(e-1)$ from the following simple argument: If $1 = 1/x_1 + \cdots + 1/x_k$, and x_k is the largest of the x_i's, say $x_k = \alpha k$ with $\alpha > 1$, then

$$1 = \sum_{j=1}^{k} \frac{1}{x_j} \ge \sum_{j=0}^{k-1} \frac{1}{x_k - j} \ge \int_{(\alpha-1)k}^{\alpha k} \frac{dt}{t} = \log\left(\frac{\alpha}{\alpha - 1}\right),$$

so that $\alpha > e/(e-1) + o(1)$ and thus $x_k \ge e(k + o(1))/(e-1)$. In [20] Greg Martin solved this question of Erdős by showing that one can express 1 as a sum of k distinct unit fractions, whose denominators are no bigger than $e(k+o(1))/(e-1)$. In fact, he proved considerably more, and showed an analogous result for unit fraction expansions for any positive rational number (not just 1).

In [12] and [15], Erdős and Graham asked similar questions about the smallest denominator when 1 is written as a sum of distinct unit fractions. For example, they asked whether it is always possible to have a representation $1 = 1/x_1 + \cdots + 1/x_k$, $x_1 < x_2 < \cdots < x_k$, where $x_1 \sim k/(e-1)$ (note: the question is incorrectly stated in [12], but is correct in [15]). Another, related question posed in the same work, is whether it is possible to have $x_k - x_1 \sim k$. Both of these questions are answered in the affirmative for infinitely many values of k in Chapter 5, where we show the more general result: If $r > 0$ is a rational number, and $N > 1$, then there exist integers $x_1, ..., x_k$, for some k, with

$$N < x_1 < x_2 < \cdots < x_k \leq \left(e^r + O_r\left(\frac{\log\log N}{\log N}\right)\right) N$$

such that

$$r = \frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_k}.$$

(note that the big-Oh error term does not depend on k directly, but does implicitly, because $k \sim (e^r - 1)N$.) This result has been accepted by *Acta Arithmetica*, but has not yet appeared. In the next chapter we give a simple argument which shows that this implies $x_k - x_1 \sim k$ and $x_1 \sim k/(e-1)$.

One consequence of this result is that there are $\sim \log N$ different representations of 1 as a sum of distinct unit fractions with denominators $\leq N$ (essentially one representation per interval of the form $[e^j, e^{j+1}]$), where no unit fraction can occur in more than one representation. A further consequence is that all of the integers $< (1 - o(1)) \log N$ can be written as a sum of unit fractions $\leq N$. We note that this result is close to best possible, since the largest integer we can so represent is $< \sum_{1 \leq n \leq N} 1/n = \log N + O(1)$. Along these lines, Erdős and Graham (see [12]) asked for an estimate of the smallest positive integer not in S_N, which is defined to be the set of all integers n expressible as $n = 1/x_1 + \cdots + 1/x_k$, for some k, where $1 \leq x_1 < \cdots < x_k \leq N$. Let s_N be the largest integer such that if $1 \leq n \leq s_N$, then $n \in S_N$; or, alternatively, s_N is the largest integer so that $s_N + 1$ is not in S_N. In

Chapter 4 we answer this question of Erdős and Graham, by proving that

$$\left[\sum_{1\le n\le N}\frac{1}{n}-\frac{9}{2}\frac{(\log\log N)^2(1+o(1))}{\log N}\right] \le s_N \le \left[\sum_{1\le n\le N}\frac{1}{n}-\frac{1}{2}\frac{(\log\log N)^2(1+o(1))}{\log N}\right],$$

which is much stronger than the bounds $(1-o(1))\log N < s_N < \log N + O(1)$ that follow from the results in Chapter 5 (though this is a much more general theorem). This result is taken from a paper which has been accepted for publication in *Mathematika.*

A common method for generating new conjectures in mathematics is to restrict or change the set of allowable values of parameters under consideration, and in the case of unit fractions, this has lead to many interesting, unsettled questions. For instance, Stein, Graham and Selfridge (see [12] and [15], p. 160) conjectured that the following modified Fibonacci algorithm always terminates:

> Suppose we are given a positive rational A/B, where B is odd. Find the largest unit fraction of the form $1/(2n+1)$ which when subtracted from A/B, leaves a non-negative remainder. Continue subtracting the largest such odd unit fraction from the successive remainders, until the remainder is itself a unit fraction.

This conjecture is still unsolved, though it is known (see [13]) that every rational number with an odd denominator can be written as a sum of distinct odd unit fractions.

One can put this type of question into the following, more general context: Given a polynomial $f(x_1,...,x_k)\in\mathbb{Z}[x_1,x_2,...,x_k]$, what can be said about the rational numbers expressible as a finite sum of distinct, positive unit fractions, where the denominators are values attained by the the polynomial f for $(x_1,...,x_k)\in\mathbb{Z}^k$?

There can be no general procedure for deciding whether any particular rational is expressible in this way, which works for every possible f, since this is equivalent to Hilbert's Tenth Problem, which was shown to be unsolvable by Y. Matiyasevich (see [21]). To see this, suppose $m/n > 0$ is a given rational number, and write it as $m/n = 1/n_1 + \cdots + 1/n_t$, where say $0 < n_1 < \cdots < n_t$ are generated by Fibonacci's algorithm. Let $g(x_1, ..., x_k)$ be any generic polynomial and set

$$f(x_1, ..., x_{k+t}) = (Cg(x_1, ..., x_k)^2 + 1)h(x_{k+1}, ..., x_{k+t}),$$

where

$$h(y_1, ..., y_t) = 1 + \sum_{j=1}^{t} (n_j - 1)y_j^{2t},$$

and C is any integer satisfying

$$C > \frac{2n}{m} \sum_{(y_1,...,y_t)\in\mathbb{Z}^t} \frac{1}{h(y_1, ..., y_t)} > 0.$$

(Note: the number of integer values $\leq x$ attained by h is at most the number of integer lattice points $(y_1, ..., y_t)$ inside the box $|y_j| < x^{1/(2t)}, j = 1, ..., t$, which is $< 2^t\sqrt{x}$. Thus, sum of reciprocals of values attained by h converges.) We claim that m/n can be written as a sum of reciprocals of distinct values attained by f if and only if there is a solution vector $(a_1, ..., a_k) \in \mathbb{Z}^k$ to $g(a_1, ..., a_k) = 0$, and deciding if such an $(a_1, ..., a_k)$ exists, for any given g, is Hilbert's Tenth Problem. Let F, G, and H be the set of distinct values attained by f, $Cg(x_1, ..., x_k)^2 + 1$, and h, respectively. Then, $F = GH$, the set of all products ab, where $a \in G$ and $b \in H$. If g is never 0 we will have

$$\begin{aligned}\sum_{j\in F} \frac{1}{j} \leq \left(\sum_{a\in G} \frac{1}{a}\right)\left(\sum_{b\in H} \frac{1}{b}\right) &\leq \left(\sum_{j=1}^{\infty} \frac{1}{Cj^2+1}\right) \sum_{(y_1,...,y_t)\in\mathbb{Z}^t} \frac{1}{h(y_1, ..., y_t)} \\ &< \frac{2}{C} \sum_{(y_1,...,y_t)\in\mathbb{Z}^t} \frac{1}{h(y_1, ..., y_t)} < \frac{m}{n}.\end{aligned}$$

If $g(a_1, ..., a_k) = 0$, for some $(a_1, ..., a_k) \in \mathbb{Z}^k$, then we have the solution

$$\frac{m}{n} = \sum_{j=1}^{t} \frac{1}{n_j} = \frac{1}{f(a_1, ..., a_k, 1, 0, 0, 0, ..., 0)} + \frac{1}{f(a_1, ..., a_k, 0, 1, 0, 0, ..., 0)} + \cdots$$
$$\cdots + \frac{1}{f(a_1, ..., a_k, 0, 0, ..., 0, 1)}.$$

Even though the case of a general polynomial f is unsolvable, it still may be possible to classify all the rationals corresponding to polynomials of sufficiently low degree, or with a small number of variables. In the case where $f(x)$ is a degree 1 polynomial[8], so that $f(x) = ax + b$, R. Graham (see [13]) showed that the corresponding rationals are all those of the form m/n, where

$$\gcd\left(\frac{n}{\gcd(n, \gcd(a, b))}, \frac{a}{\gcd(a, b)}\right) = 1.$$

In [14], Graham answered the question when $f(x) = x^k$, for all $k \geq 2$ (note: the case $k = 1$ is Fibonacci's result.), and showed that the corresponding rationals lie in certain half-open intervals. Just to take the simplest case, $f(x) = x^2$, he showed that these rationals are all those contained in the two intervals $[0, \pi^2/6 - 1)$ and $[1, \pi^2/6)$.

One can ask such questions for unit fractions with denominators in other "natural" subsets of the positive integers, such as integers with a restricted number of prime divisors, or subsets with positive lower density. [9] Indeed, Erdős and Graham showed (see [12]) that any positive rational A/B, where B is square-free, can be written as a sum of unit fractions $1/n$, where each n is the product of exactly three distinct primes, and they asked whether an analogous result holds for n's which are the product of only two distinct primes, which is still an unsettled question. In regard to subsets of the integers with positive lower density, they made many

[8]We note that all degree 1 polynomials in k variables take on the same integer values of some other degree 1 polynomial in 1 variable.

[9]A sequence $a_1 < a_2 < \cdots < a_k$ is said to have positive lower density if $\#\{a_i \leq x\}/x$ is bounded from below by some positive number for all sufficiently large x

conjectures, some of which are still unsolved. We list a few of these and related conjectures which appeared in [12].

1. Is it true that any sequence $x_1 < x_2 < \cdots < \cdots$ of positive density contains a finite subset whose sum of reciprocals equals 1?

This is still unsolved, although the methods in Chapter 6 could possibly be modified to answer it.

2. Let $A(n)$ denote any largest subset of $\{1, 2, ..., n\}$ containing no subset whose sum of reciprocals equals 1. Is $|A(n)| = n - o(n)$?

3. Suppose one partitions the integers ≥ 2 into r classes. [10] Does there always exist a finite set S belonging entirely to one of the classes, such that $\sum_{n \in S} 1/n = 1$? (note: this question also appears in [22])

In Chapter 6 we prove a general result, which shows for question 2 that $|A(n)| < cn$, for some constant $c < 1$ so that the answer in question 2 is "no", and which answers question 3 in the affirmative. Moreover, for question 3, we show that there exists such a set S, which is a subset of $[2, e^{167000r+o(r)}]$. We claim that the 167000 cannot be improved to a number less than 1, though it may be possible to improve it to 1 exactly. To see this, let $B_1, ..., B_r$ denote consecutive blocks of integers, starting with 2, where the sum of reciprocals of integers in each block is just under 1 (so, $B_1 = \{2, 3\}$, since $1/2 + 1/3 < 1$ and $1/2 + 1/3 + 1/4 > 1$), and let x be the largest integer in B_r, so that $B_1 \cup \cdots \cup B_r = [2, x] \cap \mathbb{Z}$. Now, $x > e^{r - o(r)}$, since

$$\sum_{n=2}^{x} \frac{1}{n} = \sum_{j=1}^{r} \sum_{n \in B_j} \frac{1}{n} > \sum_{j=1}^{r} \left(1 - \frac{1}{j}\right) = r - \log r - O(1).$$

Thus, $B_1, ..., B_r$ form a partition of the integers in $[2, e^{r-o(r)}]$, where no class contains a subset S with $\sum_{n \in S} 1/n = 1$.

[10] This can be rephrased in the language of colorings as follows: Suppose one r-colors the integers ≥ 2

Chapter 2

Review of Literature

We give here a review of the literature which pertains directly to the results in Chapters 4, 5, and 6.

As in the Introduction, define S_N to be the collection of all integers n which can be written as $n = 1/x_1 + \cdots + 1/x_k$, for some k, and where $1 \le x_1 < x_2 < \cdots < x_k \le N$. Define s_N to be the largest integer so that if $1 \le n \le s_N$, then $n \in S_N$. Erdős asked for an estimate for $|S_N|$ in [12], and noted that $|S_N| < \log N + O(1)$, trivially. In [33], H. Yokota proved that $|S_N| > (1/2 + o(1)) \log N$. He improved this to an asymptotic estimate, $|S_N| = \log N + O(\log\log N)$ in [35], and in his most recent paper (see [36]), he shows that if m_N denotes the largest integer in S_N, then there exist two positive constants, c_1 and c_2, such that

$$\log N + \gamma - 2 - \frac{c_1}{\log\log N} < m_N < \log N + \gamma - \frac{c_2}{\log\log N}.$$

This result contradicts the results in Chapter 4, where we obtain

$$\left[\sum_{1 \le n \le N} \frac{1}{n} - \frac{9}{2} \frac{(\log\log N)^2 (1 + o(1))}{\log N} \right] \le s_N \le \left[\sum_{1 \le n \le N} \frac{1}{n} - \frac{1}{2} \frac{(\log\log N)^2 (1 + o(1))}{\log N} \right],$$

and the reason for this is that Yokota's result is incorrect. [1]

[1] Yokota's upper bound for m_N is incorrect, and a correction recently appeared in the July 2000 issue of *J. Number Theory*. His upper bound is now the same as for s_N listed above.

In another paper (see [34]), Yokota proves that for infinitely many k, there exists an integer sequence $0 < x_1 < \cdots < x_k$, with $1 = 1/x_1 + \cdots + 1/x_k$, where

$$x_k < k(\log\log k)^3.$$

In [19], Greg Martin improves upon this, using methods similar to those in Chapter 4, by showing that any rational number r can be written as $r = 1/x_1 + \cdots + 1/x_k$, where $x_1 < x_2 < \cdots < x_k$ and $x_k = O_r(k)$. In [20], he improves this further by showing that such a representation exists, for all sufficiently large k, with

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

Furthermore, the error term is best-possible, up to a constant factor. Martin's representation for r has the property that the number of x_j's in $[x_k/e^r, x_k]$ is $\sim k$, which suggests that perhaps there is representation with

$$x_k - x_1 \sim k, \text{ and } x_1 \sim \frac{x_k}{e^r} \sim \frac{k}{e^r - 1}.$$

which were questions posed by Erdős and Graham in [12] (their questions were with $r = 1$). This is indeed the case, and is a consequence of our main result in Chapter 5, where we prove the following: There exist integers $x_1, ..., x_k$ with $r = 1/x_1 + \cdots + 1/x_k$, for some k, where

$$N < x_1 < x_2 < \cdots < x_k \leq \left(e^r + O_r\left(\frac{\log\log N}{\log N}\right)\right) N,$$

for all $N \geq 1$; moreover, the error term is best-possible. To see that this answers the questions of Erdős and Graham, let $I = [N, (e^r + O_r(\log\log N/\log N)N]$, where the big-Oh error term is the same as above. Then, we have

$$\sum_{\substack{j \in I \cap \mathbb{Z} \\ j \neq x_1, \ldots, x_k}} \frac{1}{j} = \sum_{j \in I \cap \mathbb{Z}} \frac{1}{j} - \sum_{j=1}^{k} \frac{1}{x_j} = \{r + o(1)\} - r = o(1),$$

and so $k \sim |I|$, the length of I, which means $x_k - x_1 \sim k$ as claimed.

Another question asked by Erdős and Graham in [12] and [22] is whether there exists a monochromatic solution to $\sum_{j=1}^{k} 1/x_j = 1$, for some k and any r-coloring of the integers ≥ 2. The first and only result on this question, besides those in Chapter 6, appeared in a paper by T. Brown and V. Rödl in [1]. They proved that there exists a monochromatic solution to $1/x_0 = 1/x_1 + \cdots + 1/x_k$, which was done by extending a theorem due to Rado on monochromatic solutions of homogeneous linear equations. In Chapter 6, we solve this question of Erdős and Graham and show that there is a monochromatic solution with $\{x_1, ..., x_k\} \subset [2, b^{r+o(r)}]$, for some constant b (which we show may be taken to be e^{167000}). We note that b cannot be taken to be smaller than e, since for any $\epsilon > 0$ and r sufficiently large, the integers in $[2, e^{(1-\epsilon)r}]$ can be r-colored in such a way that the sum of of reciprocals of the numbers of each color is just under 1, as was shown in the Introduction.

Let S be any subset of $\{1, 2, ..., N\}$ with $\lambda(N) = \sum_{n \in S} 1/n$ maximal such that there is no subset of S whose sum of reciprocals equals 1. Trivially, $\lambda(N) > \log\log N + O(1)$, because no subset of the primes has sum of reciprocals equal to 1 and since $\sum_{p \leq N} 1/p = \log\log N + O(1)$; also, $\lambda(N) < \sum_{j=1}^{N} 1/j = \log N + O(1)$. In an email to me, Carl Pomerance pointed out the set of all integers $n > 1$ with a prime factor $> n(1 + o(1))/\log n$ cannot contain a finite subset whose sum of reciprocals equals 1 (for a suitable $o(1)$), and so

$$\lambda(N) \gg \sum_{\substack{p \leq N/\log N \\ p \text{ prime}}} \sum_{m \leq \log p} \frac{1}{mp} \gg \sum_{\substack{p \leq N/\log N \\ p \text{ prime}}} \frac{\log\log p}{p} \gg (\log\log N)^2.$$

A consequence of our result in Chapter 6 is that $\lambda(N) < c \log N$, for some constant $c < 1$, which answers another question of Erdős and Graham mentioned in [12]. It would be nice to get a good estimate for $\lambda(N)$.

Chapter 3

Notation, Definitions, and Standard Theorems

In analytic and combinatorial number theory there are a fair number of notations and methods which are used over and over again, and this chapter is devoted to listing some of them for future reference.

3.1 Arithmetical Functions

Here are a few standard arithmetical functions used in number theory:

$$\omega(n) = \#\{p \ : \ p|n, \ p \text{ prime}\} = \sum_{\substack{p|n \\ p \text{ prime}}} 1,$$

$$\Omega(n) = \#\{p^a \ : \ p^a|n, \ p \text{ prime}\},$$

$$\mu(n) = \begin{cases} 0, & \text{if } p^2|n, \text{ for some prime } p; \\ (-1)^{\omega(n)}, & \text{otherwise.} \end{cases}$$

$$\tau(n) = \#\{d \ : \ d|n\} = \sum_{d|n} 1$$

$$\sigma(n) = \sum_{d|n} d$$

$$\varphi(n) = \#\{m \ : \ 1 \le m \le n, \ \gcd(m,n) = 1\} = n \prod_{\substack{p|n \\ p \text{ prime}}} \left(1 - \frac{1}{p}\right),$$

$$\Lambda(n) = \begin{cases} \log p, & \text{if } n = p^a, \ p \text{ prime} \\ 0, & \text{otherwise,} \end{cases}$$

$$\psi(x) = \sum_{n \le x} \Lambda(n) = \log\left(\text{lcm}\{2, 3, ..., [x]\}\right),$$

$$\pi(x) = \#\{p \le x \ : \ p \text{ prime}\}.$$

3.2 Big-Oh, Little-Oh, and Vinogradov Notation

Given two functions $f(x)$ and $g(x)$, we say that $f(x) = O(g(x))$, which is read "$f(x)$ is big-Oh of $g(x)$", or in Vinogadov's notation, $f(x) \ll g(x)$ (or, $g(x) \gg f(x)$), if there exists a constant C so that $|f(x)| < C|g(x)|$, whenever $x > X_0$. If f and g happen to depend on several parameters, then we use subscripted notation to indicate which parameters the implied constant C depends on; for instance, if we write $f(x) = O_a(g(x))$, or $f(x) \ll_a g(x)$, where $f(x)$ and $g(x)$ depend on some invisible parameter a, we mean that there exists some constant $C(a)$, depending only on a, such that $|f(x)| < C(a)|g(x)|$, whenever $x > X_0(a)$.

If $g(x) = O(f(x))$ and $f(x) = O(g(x))$, then we use the notation $f(x) \asymp g(x)$, and if $\lim_{x\to\infty} f(x)/g(x) = 1$, then we say that $f(x)$ is asymptotic to $g(x)$, and we denote this by $f(x) \sim g(x)$. We note that if $f(x) \sim g(x)$, then $f(x) \asymp g(x)$, but the converse is not true.

We say that $f(x) = o(g(x))$, which is read "$f(x)$ is little-Oh of $g(x)$" if $\lim_{x\to\infty} f(x)/g(x) = 0$, and we commonly use the notation $o(1)$ to represent some generic function which tends to 0 as some implicit parameter tends to infinity (and $O(1)$ represents a function which remains bounded).

3.3 Partial Summation

The method of partial summation is used frequently in analytic number theory, and is a method for estimating the sum $S(x) = \sum_{1\le n\le x} a_n f(n)$, if one has an estimate for $T(t) = \sum_{1\le n\le t} a_n$, for $t \le x$, where a_n is some arbitrary sequence and $f(x)$ is some differentiable function. The way the method works is by the following formula

$$S(x) = f(x)T(x) - \int_1^x T(t)f'(t)\, dt$$

An example of how this formula is used is as follows. Suppose we want to estimate the sum

$$S(x) = \sum_{\substack{p \leq x \\ p \text{ prime}}} \frac{\log p}{p},$$

if we know that

$$T(t) = \sum_{\substack{p \leq t \\ p \text{ prime}}} \log p = t + O\left(\frac{t}{\exp((\log t)^{3/5-o(1)})}\right).$$

Then, we have that $f(t) = 1/t$, which is differentiable (when $t \neq 0$), and so by the partial summation formula

$$\begin{aligned} S(x) &= \frac{1}{x}\left(x + O\left(\frac{x}{\exp((\log x)^{3/5-o(1)})}\right)\right) \\ &\qquad + \int_1^x \left(t + O\left(\frac{t}{\exp((\log t)^{3/5-o(1)})}\right)\right)\frac{1}{t^2}\,dt \\ &= O(1) + \int_1^x \frac{dt}{t} + \int_1^x O\left(\frac{1}{t\exp((\log t)^{3/5-o(1)})}\right)dt \\ &= O(1) + \log x + O\left(\int_1^x \frac{1}{t\ \exp((\log t)^{3/5-o(1)})}\,dt\right) = \log x + O(1). \end{aligned}$$

3.4 Mertens's Theorem

Mertens's Theorem states that

$$\prod_{p \leq x}\left(1 - \frac{1}{p}\right) \sim \frac{e^{-\gamma}}{\log x},$$

where $\gamma = 0.57721566...$ is Euler's constant. Another form of this result is as follows:

$$\sum_{\substack{p \leq x \\ p \text{ prime}}} \frac{1}{p} = \log\log x + A + O\left(\frac{1}{\log x}\right),$$

where

$$A = \gamma + \sum_{p \text{ prime}} \left\{\log\left(1 - \frac{1}{p}\right) + \frac{1}{p}\right\}.$$

3.5 The Prime Number Theorem

The Prime Number Theorem with good error term (see [27]), states that

$$\pi(x) \;=\; \mathrm{li}(x) + O\left(\frac{x}{\exp((\log x)^{3/5-o(1)})}\right),$$

where

$$\mathrm{li}(x) \;=\; \int_2^x \frac{dt}{\log t} \sim \frac{x}{\log x},$$

and another frequently used form of it is:

$$\psi(x) \;=\; x + O\left(\frac{x}{\exp((\log x)^{3/5-o(1)})}\right).$$

We note that these two forms of the Prime Number Theorem are interchangeable through the use of partial summation.

3.6 Exponential Sums

An exponential sum is a function $f(t)$ which can be expressed as

$$f(t) = \sum_{n\le N} a_n \; e^{2\pi int}.$$

We use the notation $e(t) = e^{2\pi it}$ to make it easier to read, so that

$$f(t) = \sum_{n\le N} a_n e(nt).$$

The reason exponential sums are so useful is because of the following two formulae:

$$\sum_{n\equiv a \pmod b} a_n \;=\; \frac{1}{b}\sum_{j=0}^{b-1} e(-aj/b)f(j/b),$$

and

$$a_n \;=\; \int_0^1 e(-nt)f(t)\;dt.$$

Both of these formulae can be proved using the following, more basic results for integers n:

$$\int_0^1 e(nt)\,dt \;=\; \begin{cases} 1, & \text{if } n = 0 \\ 0, & \text{if } n \neq 0, \end{cases} \qquad \text{and} \quad \frac{1}{b}\sum_{j=0}^{b-1} e(jn/b) \;=\; \begin{cases} 1, & \text{if } n \equiv 0 \pmod{d} \\ 0, & \text{if } n \not\equiv 0 \pmod{d}. \end{cases}$$

Chapter 4

On Some Questions of Erdős and Graham about Egyptian Fractions

4.1 Introduction

Define S_N to be the set of all positive integers m which can be expressed as

$$m = \frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_k},$$

where k is variable and the x_i's are integers with $1 \leq x_1 < x_2 < \cdots < x_k \leq N$. In [12] and [15], Erdős and Graham asked the following questions.

1. What is the smallest natural number not in S_N?
2. How many numbers are in S_N?

Recently, in [35], Yokota showed that $\{n \ : \ 1 \leq n \leq \log N - 5 \log\log N\} \subseteq S_N$, thus giving the correct asymptotic to the first two of these questions. In this chapter we prove the following results.

Main Theorem *Define s_N to be the largest integer such that, whenever $1 \leq n \leq s_N$, $n \in \mathbb{Z}$, there exist integers $1 \leq x_1 < x_2 < \cdots < x_k \leq N$, for some k, such that*

$$n = \frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_k}.$$

Then

$$\left[\sum_{1 \leq n \leq N} \frac{1}{n} - \frac{9}{2} \frac{(\log\log N)^2 (1 + o(1))}{\log N} \right] \leq s_N \leq \left[\sum_{1 \leq n \leq N} \frac{1}{n} - \frac{1}{2} \frac{(\log\log N)^2 (1 + o(1))}{\log N} \right].$$

Corollary 4.1 *Let* $\gamma = \lim_{N\to\infty} \left\{\sum_{1\le n\le N} 1/n - \log N\right\}$ *be Euler's constant. For a given positive integer* n *there are integers*

$$1 \le x_1 < x_2 < \cdots < x_k \le e^{n-\gamma}\left\{1 + \left(\frac{9}{2} + o(1)\right)\frac{\log^2 n}{n}\right\},$$

for some k*, such that*

$$n = \frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_k}.$$

To answer these questions of Erdős and Graham above, let $\sum_{1\le n\le N} 1/n = m + \delta$, where $m = m(N)$ is the integer part of $\sum_{1\le n\le N} 1/n$, and $\delta = \delta(N)$ is the fractional part.

We have trivially that $S_N \subseteq \{1, 2, ..., m\}$ and our main theorem tells us that when N is sufficiently large, $\{1, 2, ..., m-1\} \subseteq S_N$. Moreover, if $\delta > (\frac{9}{2} + o(1))(\log\log N)^2/\log N$ then $m \in S_N$ so that $S_N = \{1, 2, ..., m\}$, and if $\delta < (\frac{1}{2} + o(1))(\log\log N)^2/\log N$ then $m \notin S_N$ so that $S_N = \{1, 2, ..., m-1\}$. Let $D_N = (\frac{1}{2} + o(1))(\log\log N)^2/\log N$. We believe that the upper bound in the Main Theorem is the truth, which if true would say that for N sufficiently large,

$$S_N = \begin{cases} \{1, 2, ..., m\}, & \text{if } \delta > D_N \\ \{1, 2, ..., m-1\}, & \text{if } \delta < D_N. \end{cases}$$

To prove the Main Theorem we will need to introduce some notation. For any given prime power p^a and any integer $N \ge 1$, define $S(p^a, N)$ to be the set of integers $\le N$ whose prime power divisors are $< p^a$. Define

$$f(p^a, N) := \max_{1\le l\le p-1} \min\left\{\frac{1}{y_1} + \cdots + \frac{1}{y_k} : \frac{1}{y_1} + \frac{1}{y_2} + \cdots + \frac{1}{y_k} \equiv l \pmod p, \right.$$
$$\left. 1 \le y_1 < \cdots < y_k \le N \text{ each } y_i \in S(p^a, N)\right\}$$

(let $f = \infty$ if such a 'maximum' does not exist, which will happen if there does not exist an k-tuple $(y_1, ..., y_k)$ for some $l \pmod p$). For $c > 0$ let

$$F(N, c) = \sum_{p^a \le N/\log^c N} \frac{f(p^a, N/p^a)}{p^a} + \sum\left\{\frac{1}{mp^a} : N/\log^c N < p^a \le N,\ mp^a \le N\right\}.$$

The idea of the proof of the Main Theorem is as follows: we start with the full sum $\sum_{1\le n\le N} 1/n$ and then try to remove as few terms as we can so that the sum of the remaining terms is an integer. We want the contribution of those terms we remove to be as small as possible and Proposition 4.1 below tells us that this contribution need be no bigger than $F(N,c)$, for any $c>0$.

Proposition 4.1 *For any given integer N there exists a subset T of the integers $\le N$ such that $S_0 = S_0(N) = \sum_{t\in T} 1/t$ is an integer satisfying*

$$0 < \sum_{n\le N} \frac{1}{n} - S_0(N) \le F(N,c),$$

for all $c \ge 0$.

To see why Proposition 4.1 is true, we first remove all of those terms in the full sum where n has a prime power factor bigger than $N/\log^c N$, which accounts for the second summand in the definition of $F(N,c)$ above. Call the sum of the remaining terms $S = u/v$, where $gcd(u,v) = 1$. We observe that if $p^a | v$ then $p^a \le N/\log^c N$. Let $q^b \le N/\log^c N$ be the largest prime power dividing v. We try to find numbers $1 \le y_1 < y_2 < \cdots < y_k \le N/q^b$, where:

1. $q \nmid y_i$ for $i = 1,2,...,k$.
2. All of the prime power factors of the y_i's are $< q^b$; and
3. $$q^b S - \tfrac{1}{y_1} - \tfrac{1}{y_2} - \cdots - \tfrac{1}{y_k} \equiv 0 \pmod q.$$

(Notice that $q^b S = q^b u/v$ makes sense modulo q because $q^b || v$.) From the definition of the function f we have that if $f(q^b, N/q^b) \ne \infty$ then there are such integers y_i; moreover, there is a choice with

$$\frac{1}{y_1} + \frac{1}{y_2} + \cdots + \frac{1}{y_k} \le f(q^b, N/q^b).$$

Let us just assume for the moment that $f(q^b, N/q^b) \ne \infty$ and let $1 \le y_1 < y_2 < \cdots < y_k \le N/q^b$ be any choice satisfying

$$\frac{1}{y_1} + \cdots + \frac{1}{y_k} \le f(q^b, N/q^b).$$

We make the following three deductions.

1. Each of the numbers

$$\frac{1}{q^b y_1}, \frac{1}{q^b y_2}, \ldots, \frac{1}{q^b y_k}$$

are terms in the sum S.

2. If we remove these terms from S and call the new sum S' so that

$$\frac{u'}{v'} = S' = S - \frac{1}{q^b y_1} - \cdots - \frac{1}{q^b y_k}, \quad \gcd(u', v') = 1,$$

then the largest prime power dividing v' is strictly smaller than q^b.

3. This new sum S' satisfies $S - S' < f(q^b, N/q^b)/q^b$.

Now let r^s be the largest prime power dividing v'. We subtract terms from sum $S'' = u''/v''$, $\gcd(u'', v'') = 1$, where the largest prime power dividing v'' is strictly smaller than $r^s < q^b$. This new sum S'' satisfies

$$S - S'' = (S - S') + (S' - S'') \leq \frac{f(q^b, N/q^b)}{q^b} + \frac{f(r^s, N/r^s)}{r^s}.$$

If we continue subtracting terms in this manner we eventually get down to a sum S_0 where

$$S - S_0 \leq \sum_{2 \leq p^a \leq N/\log^c N} \frac{f(p^a, N/p^a)}{p^a},$$

and S_0 is an integer. Proposition 4.1 now follows since

$$\sum_{1 \leq n \leq N} 1/n - S_0 \leq F(N, c).$$

We will obtain explicit bounds on $F(N, c)$ by proving the following inequality for $f(p^a, N/p^a)$:

$$f(p^a, N/p^a) \leq \begin{cases} (p-1)p^a/\mathrm{lcm}\{2, 3, 4, ..., p^a\}, & \text{if } p^a < \frac{1}{4}\log N \\ 20/(p^a - 1), & \text{if } \frac{1}{4}\log N < p^a \leq \sqrt{N} \\ 4/\log^{\epsilon/3} N, & \text{if } \sqrt{N} < p^a \leq N/(\log^{3+\epsilon} p^a). \end{cases}$$

To prove the first case, when $p^a < \frac{1}{4}\log N$, we will show that

$$\left\{\frac{\mathrm{lcm}\{2,3,...,p^a\}}{p^a u} \;:\; 1 \le u \le (p-1)\right\} \subseteq S(p^a, N),$$

and moreover that this set has a member in each residue class $\not\equiv 0 \pmod{p}$. Our bound $f(p^a, N/p^a) \le (p-1)p^a/\mathrm{lcm}\{2,3,...,p^a\}$ then follows. Using the following identity we will have that the contribution of such small prime powers to $F(N,c)$ is < 1.

Lemma 4.1 *For q prime and positive integer b,*

$$\sum\left\{\frac{p-1}{\mathrm{lcm}\{2,3,...,p^a\}} \;:\; 2 \le p^a \le q^b, p \text{ prime}\right\} = \frac{\mathrm{lcm}\{2,3,...,q^b\}-1}{\mathrm{lcm}\{2,3,...,q^b\}}.$$

The bound on $f(p^a, N/p^a)$ where $\log N/4 < p^a \le \sqrt{N}$ comes directly from the following lemma:

Lemma 4.2 *If $p \neq 2$ then $f(p^a, p^a) < 20/(p^a - 1)$ for $p^a > 3$.*

From this lemma we will show that the contribution of prime powers p^a with $\log N/4 < p^a \le \sqrt{N}$ to $F(N,c)$ is $O(1/\log N)$.

Finally, the bound on $f(p^a, N/p^a)$ where $\sqrt{N} < p^a < N/\log^{3+\epsilon} N$ follows from the following Proposition and its corollary.

Proposition 4.2 *Suppose $\epsilon > 0$ is given. There exists a number $\hat{N}_\epsilon$ such that whenever $n > \hat{N}_\epsilon$ and $k > \log^{3+2\epsilon} n$, for any set of k distinct primes $2 \le p_1 < p_2 < \cdots < p_k < \log^{3+3\epsilon} n$ which do not divide n, and any residue class $l \pmod{n}$, there is a subset*

$$\{q_1, q_2, ..., q_t\} \subseteq \{p_1, p_2, ..., p_k\}$$

such that

$$\frac{1}{q_1} + \frac{1}{q_2} + \cdots + \frac{1}{q_t} \equiv l \pmod{n}.$$

Corollary 4.2 *Let $\delta > 0$ be given. There exists a constant M_δ so that, when $M_\delta < p^a \leq N/\log^{3+\delta} N$,*

$$f(p^a, N) < \frac{4}{\log^{\delta/3} p^a}.$$

We will show that the contribution of such prime powers to $F(N, 3+\epsilon)$ is $O(1/\log N)$, and that the contribution of the prime powers p^a with $N/\log^{3+\epsilon} N \leq p^a \leq N$ to $F(N, 3+\epsilon)$ is $(\frac{9}{2} + o(1))(\log\log N)^2/\log N$, and so we arrive at

Proposition 4.3 *For all $\epsilon > 0$ we have*

$$F(N, 3+\epsilon) < 1 + \left(\frac{(3+\epsilon)^2}{2} + o(1)\right)\frac{(\log\log N)^2}{\log N}.$$

With Propositions 4.1 and 4.3, and the fact that every integer can be written as a sum of unit fractions (see [35]), we prove the bound

$$s_N \geq \left[\sum_{1\leq n\leq N}\frac{1}{n} - \left(\frac{9}{2} + o(1)\right)\frac{(\log\log N)^2}{\log N}\right],$$

as claimed in the Main Theorem. To get the bound

$$s_N \leq \left[\sum_{1\leq n\leq N}\frac{1}{n} - \left(\frac{1}{2} + o(1)\right)\frac{(\log\log N)^2}{\log N}\right]$$

we show that if $1 \leq x_1 < x_2 < \cdots < x_k \leq N$ and

$$\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_k}$$

is an integer, then none of the x_i's can be divisible by a prime p with $N\log\log N/\log N < p \leq N$. From this and a technical lemma it follows that

$$\sum_{1\leq n\leq N}\frac{1}{n} - \frac{1}{x_1} - \cdots - \frac{1}{x_k} \geq \sum_{\substack{N\log\log N/\log N < p\leq N\\ mp\leq N}}\frac{1}{mp} > \left(\frac{1}{2} + o(1)\right)\frac{(\log\log N)^2}{\log N}.$$

4.2 Proofs and Technical Lemmas

Proof of Lemma 4.1. The lemma holds for $q^b = 2$, since

$$\sum_{2 \le p^a \le 2} \frac{p-1}{\operatorname{lcm}\{2, ..., p^a\}} = \frac{1}{2}.$$

Assume, for proof by mathematical induction, we have shown that the theorem holds for all prime powers q^b, where $2 \le q^b < r^s$, where r^s is some prime power. We observe that $\operatorname{lcm}\{n \ : \ 2 \le n \le t\} = \operatorname{lcm}\{p^a \ : \ 2 \le p^a \le t, \ p \text{ prime}\}$. Using this and the induction hypothesis we have

$$\begin{aligned}
\sum_{2 \le p^a \le r^s} & \frac{p-1}{\operatorname{lcm}\{2,3, ..., p^a\}} \\
&= \sum_{2 \le p^a < r^s} \frac{p-1}{\operatorname{lcm}\{2,3,...,p^a\}} + \frac{r-1}{\operatorname{lcm}\{2,3,...,r^s\}} \\
&= \frac{\operatorname{lcm}\{2,3,...,r^s-2, r^s-1\} - 1}{\operatorname{lcm}\{2,3,...,r^s-2,r^s-1\}} + \frac{r-1}{\operatorname{lcm}\{2,3,...,r^s\}} \\
&= \frac{r \cdot \operatorname{lcm}\{2,3,...,r^s-2,r^s-1\} - r}{\operatorname{lcm}\{2,3,...,r^s\}} + \frac{r-1}{\operatorname{lcm}\{2,3,...,r^s\}} \\
&= \frac{\operatorname{lcm}\{2,3,...,r^s\} - 1}{\operatorname{lcm}\{2,3,...,r^s\}}, \qquad (4.1)
\end{aligned}$$

and so the theorem follows by mathematical induction.

Proof of Lemma 4.2. Suppose l is any integer where $1 \le l \le (p-1)$. The number of pairs (y_1, y_2) such that $1 \le y_1 < y_2 \le p^a - 1$, $p \nmid y_1 y_2$ and

$$\frac{1}{y_1} + \frac{1}{y_2} \equiv l \pmod{p} \qquad (4.2)$$

is $\frac{1}{2}((p-2)p^{2a-2} - 1) > 1$ for $p^a > 3$. Now, one of these pairs must have $y_1 > p^a/10$, since the number of pairs (y_1, y_2) with $y_1 < p^a/10$ satisfying (4.2) is less than

$$\frac{p^{2a-1}}{10} < \frac{1}{2}\left((p-2)p^{2a-2} - 1\right),$$

whenever $p > 2$. For this pair, we will have

$$\frac{1}{y_1} + \frac{1}{y_2} < \frac{2}{y_1} < \frac{20}{p^a}.$$

Since l was arbitrary, it follows that

$$f(p^a,p^a) < \frac{20}{p^a}.$$

Lemma 4.3 *Let $g(N)$ be the largest prime power such that* $\mathrm{lcm}\{2,3,4,...,g(N)\} \leq N$. *When $g(N) \geq 2$ we have that $g(N) > \frac{1}{4}\log N$.*

Proof of Lemma 4.3. Let $h(N)$ be the next prime power after $g(N)$. Since $\pi(N) < 2N/\log N$ when $N \geq 2$ (see [27]) we have when $h(N) \geq 2$ that

$$N < \mathrm{lcm}\{2,3,...,h(N)\} < h(N)^{\pi(h(N))} < h(N)^{2h(N)/\log h(N)} = e^{2h(N)}.$$

By Bertrand's postulate that for $n \geq 2$ there is always a prime p with $n < p < 2n$ we must have $\log N < 2h(N) < 4g(N)$ when $h(N) \geq 2$ and so $g(N) > \frac{1}{4}\log N$ when $g(N) \geq 2$.

Lemma 4.4

$$\sum_{\substack{N/\log^c N < p \leq N \\ mp^a \geq N}} \frac{1}{mp} = \frac{c^2}{2}\frac{(\log\log N)^2}{\log N} + O\left(\frac{\log\log N}{\log N}\right).$$

Notice that since

$$\sum_{\substack{a\geq 2 \\ N/\log^c N<p^a\leq N \\ mp^a\leq N}} \frac{1}{mp^a} \ll \sum_{\substack{a\geq 2 \\ N/\log^c N<p^a\leq N \\ mp^a\leq N}} \frac{1}{p^a}\sum_{m\leq \log^c N}\frac{1}{m} \ll \frac{\log\log N}{N/\log^c N}\sum_{\substack{p^a\leq N \\ a\geq 2}} 1 \ll \frac{\log^c N}{\sqrt{N}},$$

Lemma 4.4 is also true if we replace the sum over primes by a sum over prime powers between $N/\log^c N$ and N.

Proof of Lemma 4.4. Using the estimate $\sum_{p\leq N} 1/p = \log\log N + B + o(1/\log N)$ (see [27]), we have for any $N/\log^c N < t \leq N$ that

$$\begin{aligned}\sum_{t/e<p<t}\frac{1}{p} &= \log\log t - \log\log\frac{t}{e} + o\left(\frac{1}{\log t}\right) = \log\left(\frac{\log t}{\log t - 1}\right) + o\left(\frac{1}{\log t}\right) \\ &= \frac{1+o(1)}{\log t} = \frac{1+o(1)}{\log N}.\end{aligned}$$

Then,

$$\begin{aligned}
\sum_{\substack{N/\log^c N<p\le N\\ mp\le N}} \frac{1}{mp} &= \sum_{N/\log^c N<p\le N} \frac{1}{p}\sum_{m\le N/p}\frac{1}{m} = \sum_{N/\log^c N<p\le N} \frac{\log(N/p)+O(1)}{p} \\
&= \left(\sum_{1\le j\le [c\log\log N]} \ \sum_{N/e^j<p\le N/e^{j-1}} \frac{j+O(1)}{p}\right) \\
&\qquad +O\left(\sum_{N/\log^c N<p\le eN/\log^c N} \frac{c\log\log N}{p}\right) \\
&= \sum_{1\le j\le [c\log\log N]} \frac{j+O(1)}{\log N} + O\left(\frac{\log\log N}{\log N}\right) \\
&= \frac{c^2}{2}\frac{(\log\log N)^2}{\log N} + O\left(\frac{\log\log N}{\log N}\right)
\end{aligned}$$

4.3 Proof of Proposition 4.1

Let $2=q_1<3=q_2<4=q_3<\ldots$ be the sequence of prime powers. Let

$$\frac{u}{v} = \sum_{2\le n\le N}\frac{1}{n} - \sum\left\{\frac{1}{n} : n\le N,\ n=mp^a,\ p^a>N/\log^c N\right\}.$$

Choose r so that q_r is the largest prime power dividing u/v (notice that $q_r \le N/\log^c N$) and let $u_r/v_r = u/v$. Define $T_r := \{n\le N : p^a | n \Rightarrow p^a \le N/\log^c N\}$ so that $u_r/v_r = \sum_{t\in T_r} 1/t$. We shall recursively define u_j/v_j and T_j, for $j = r-1, r-2, \ldots, 0$ where $u_j/v_j = \sum_{t\in T_j} 1/t$, $T_j \subseteq T_{j+1}$, and so that

(i) $\{n\le N\ :\ p^a|n \Rightarrow p^a\le q_j\} \subseteq T_j$, and

(ii) $p^a | v_j \Rightarrow p^a \le q_j$,

where we take $q_0=1$. Then we take $T=T_0$ in the Proposition since (ii) implies $v_0=1$ so that u_0/v_0 is an integer.

If q_j does not divide v_j, let $T_{j-1}=T_j$ and $u_{j-1}/v_{j-1}=u_j/v_j$; otherwise, assume q_j divides v_j and suppose q_j is some power of the prime p. Let $l \equiv q_j u_j/v_j \pmod p$

and select, if we can, integers $1 \leq y_1 < y_2 < \cdots < y_k \leq N/q_j$, each belonging to $S(q_j, N/q_j)$, so that

$$\frac{1}{y_1} + \frac{1}{y_2} + \cdots + \frac{1}{y_k} \equiv l \pmod{p}.$$

Then let $S_j := \{q_j y_1, q_j y_2, ..., q_j y_k\}$. Note that, by (i), $S_j \subseteq T_j$. Define $T_{j-1} := T_j \backslash S_j$. Thus

$$\frac{u_{j-1}}{v_{j-1}} = \sum_{t \in T_{j-1}} \frac{1}{t} = \frac{u_j}{v_j} - \sum_{s \in S_j} \frac{1}{s}.$$

We see immediately that (i) above is satisfied. Now

$$q_j \left(\frac{u_j}{v_j} - \sum_{s \in S_j} \frac{1}{s} \right) = \frac{q_j u_j}{v_j} - \left(\frac{1}{y_1} + \cdots + \frac{1}{y_k} \right) \equiv 0 \pmod{p}.$$

Since $q_j \nmid v_{j-1}$ we have that (ii) is satisfied. Finally note that, by definition

$$\sum_{s \in S_j} \frac{1}{s} \leq \frac{f(q_j, N/q_j)}{q_j},$$

and so the Proposition follows.

4.4 Proof of Proposition 4.2 and its Corollary

Proof of Proposition 4.2. Suppose that b is coprime to n, and let $r_n(a/b)$ denote the least residue of $ab^{-1} \pmod{n}$ in absolute value. The number of subsets of $\{p_1, ..., p_k\}$ whose sum of reciprocals is $\equiv l \pmod{n}$ is then given by

$$S_l \ := \ \frac{1}{n} \sum_{h=0}^{n-1} e\left(\frac{-hl}{n} \right) \prod_{j=1}^{k} \left(1 + e\left(\frac{r_n(h/p_j)}{n} \right) \right),$$

where $e(x)$ is defined to be $e^{2\pi i x}$. Define

$$P(h) \ := \ \prod_{j=1}^{k} \left(1 + e\left(\frac{r_n(h/p_j)}{n} \right) \right).$$

We will show that

$$|P(h)| < \frac{2^k}{n}, \tag{4.3}$$

when $h \neq 0$ and when n is sufficiently large. It will then follow that

$$|S_l| = \left|\frac{1}{n}\sum_{h=0}^{n-1} P(h)\right| > \frac{1}{n}\left\{2^k - \sum_{h=1}^{n-1}\frac{2^k}{n}\right\} = \frac{2^k}{n^2} > 0,$$

and thus there is at least one subset of $\{p_1, ..., p_k\}$ with the desired property.

To prove (4.3) we note that

$$\begin{aligned} |P(h)| &= \left|\prod_{j=1}^{k} e\left(\frac{r_n(h/p_j)}{2n}\right)\left\{e\left(\frac{-r_n(h/p_j)}{2n}\right) + e\left(\frac{r_n(h/p_j)}{2n}\right)\right\}\right| \\ &= 2^k\left|\prod_{j=1}^{k}\cos\left(\pi\frac{r_n(h/p_j)}{n}\right)\right|. \end{aligned} \tag{4.4}$$

We may write

$$r_n(h/p_j) = \frac{s_j n + h}{p_j},$$

where $0 \leq h \leq (n-1)$ and s_j is an integer satisfying $-[\frac{1}{2}p_j] < s_j \leq [\frac{1}{2}p_j]$. Define $L(N) := \log^{2+2\epsilon} N + 1$. We will now show that when n is sufficiently large at least $\frac{1}{2}k$ of the s_j's have the property that $|s_j| > L(n)$. Since, if we suppose there are infinitely many n where at least $\frac{1}{2}k$ of the s_j's satisfy $|s_j| \leq L(n)$ then, by the pigeonhole principle, there is a number m with $|m| \leq L(n)$ such that $s_j = m$ for at least

$$\frac{k/2}{2L(n)+1} > \frac{\log^{3+2\epsilon} n}{4\log^{2+2\epsilon} n + 6} \gg \log n$$

of the primes p_j dividing $mn + h$ when n is sufficiently large. However, this is impossible for large n since $|mn + h| < n(L(n)+1) < n^2$ has $o(\log n)$ distinct prime factors. Thus when n is sufficiently large at least $\frac{1}{2}k$ of the s_j's satisfy $|s_j| > L(n)$.

It follows that, when n is sufficiently large, at least $\frac{1}{2}k$ of the p_j's satisfy

$$|r_n(h/p_j)| = \left|\frac{s_j n + h}{p_j}\right| > \left|\frac{(s_j - 1)n}{p_j}\right| > \frac{n}{\log^{1+\epsilon} n}.$$

We have for such primes p_j that, when n is sufficiently large,

$$\begin{aligned} \left|\cos\left(\frac{\pi r_n(h/p_j)}{n}\right)\right| &= \left|1 - \frac{1}{2}\left(\frac{\pi r_n(h/p_j)}{n}\right)^2 + O\left(\left(\frac{r_n(h/p_j)}{n}\right)^4\right)\right| \\ &< 1 - \frac{\pi^2}{2\log^{2+2\epsilon} n} + O\left(\frac{1}{\log^{4+4\epsilon} n}\right), \end{aligned}$$

and so, from (4.4), since $k > \log^{3+2\epsilon} n$ we have that

$$|P(h)| < 2^k \left(1 - \frac{\pi^2}{2\log^{2+2\epsilon} n} + O\left(\frac{1}{\log^{4+4\epsilon} n}\right)\right)^{k/2} \ll 2^k e^{-\frac{\pi^2 \log n}{4}} = o\left(\frac{2^k}{n}\right),$$

which was just what we needed to show in order to prove our Proposition.

Proof of Corollary. Let $\hat{N}_{\delta/3}$ be as in Proposition 4.2, and let $\hat{N}'_\delta$ be the smallest number so that when $N \geq \hat{N}'_\delta$ there are at least $\log^{3+\frac{2}{3}\delta} N + 1$ primes q with $\frac{1}{2}\log^{3+\delta} N \leq q \leq \log^{3+\delta} N$ where $q \neq p$. Suppose l is any number where $0 \leq l \leq (p-1)$ and p^a is any prime power where

$$N/\log^{3+\delta} N \geq p^a > M_\delta := \max\{\hat{N}_{\delta/3}, \hat{N}'_\delta\}.$$

By Proposition 4.2 there are primes q_1, q_2, ..., q_t with

$$\frac{1}{2}\log^{3+\delta} N \leq q_1 < q_2 < \cdots < q_t \leq \log^{3+\delta} N$$

and $t \leq log^{3+\frac{2}{3}\delta} N + 1$ such that

$$\frac{1}{q_1} + \frac{1}{q_2} + \cdots + \frac{1}{q_t} \equiv l \pmod{p}.$$

Also

$$\frac{1}{q_1} + \cdots + \frac{1}{q_t} < \frac{t}{\frac{1}{2}\log^{3+\delta} N} \leq \frac{2(\log^{3+\frac{2}{3}\delta} N + 1)}{\log^{3+\delta} N} < \frac{4}{\log^{\delta/3} N}.$$

Since this bound holds for all l with $0 \leq l \leq (p-1)$ the Corollary follows.

4.5 Proof of Proposition 4.3

Let $g(N)$ be as in Lemma 4.3. We may then write

$$F(N, 3+\epsilon) = A(N) + B(N) + C(N) + D(N) + E(N),$$

where

$$\begin{aligned}
A(N) &= \sum_{2\le p^a\le g(N)} \frac{f(p^a,N/p^a)}{p^a},\\
B(N) &= \sum_{g(N)<p^a\le\sqrt{N}} \frac{f(p^a,N/p^a)}{p^a},\\
C(N) &= \sum_{\sqrt{N}<p^a\le N/\log^6 N} \frac{f(p^a,N/p^a)}{p^a},\\
D(N) &= \sum_{N/\log^6 N<p^a\le N/\log^{3+\epsilon} N} \frac{f(p^a,N/p^a)}{p^a}, \text{ and}\\
E(N) &= \sum\left\{\frac{1}{mp^a} : N/\log^{3+\epsilon} N < p^a \le N,\ mp^a\le N\right\}
\end{aligned}$$

From Lemma 4.4 we have that $E(N) = (\frac{1}{2}+o(1))(3+\epsilon)^2(\log\log N)^2/\log N$. We will show that $A(N) < 1$ and that $B(N)$, $C(N)$ and $D(N)$ are each $O(1/\log N)$ and so the Proposition will follow.

For each prime power p^a define

$$U_{p^a} := \left\{\frac{\operatorname{lcm}\{2,3,...,p^a\}}{p^a u} : 1\le u\le (p-1)\right\}.$$

We have that $U_{p^a}\subset S(p^a,N/p^a)$, for $p^a\le g(N)$. Also, for each l with $1\le l\le(p-1)$, there is an element $y\in U_{p^a}$ such that $1/y\equiv l \pmod p$. Thus for $p^a\le g(N)$ we have

$$f(p^a,N/p^a)\le\frac{1}{\min(U_{p^a})}=\frac{p^a(p-1)}{\operatorname{lcm}\{2,3,...,p^a\}}.$$

From this and Lemma 4.1 it follows that

$$\begin{aligned}
A(N) &= \sum_{2\le p^a\le g(N)}\frac{f(p^a,N/p^a)}{p^a}\\
&\le \sum_{2\le p^a\le g(N)}\frac{p-1}{\operatorname{lcm}\{2,3,...,p^a\}} = \frac{\operatorname{lcm}\{2,3,...,g(N)\}-1}{\operatorname{lcm}\{2,3,...,g(N)\}}.
\end{aligned}$$

Thus $A(N) < 1$.

By Lemma 4.2, if $p \neq 2$ and $p^a \leq \sqrt{N}$ then $f(p^a, N/p^a) \leq f(p^a, p^a) < 20/p^a$. Thus by this and Lemma 4.3 we have for $g(N) \geq 2$ that

$$\begin{aligned} B(N) &= \sum_{g(N)<p^a\leq\sqrt{N}} \frac{f(p^a, N/p^a)}{p^a} < \sum_{\frac{1}{4}\log N<p^a\leq\sqrt{N}} \frac{f(p^a, N/p^a)}{p^a} \\ &\leq \sum_{2^a>\frac{1}{4}\log N} \frac{1}{2^a} + \sum_{\substack{\frac{1}{4}\log N<p^a\leq\sqrt{N} \\ p \ odd}} \frac{20}{p^{2a}} < \frac{4}{\log N} + 20 \sum_{n>\frac{1}{4}\log N} \frac{1}{n^2} \\ &= O\left(\frac{1}{\log N}\right). \end{aligned}$$

By the Corollary to Proposition 4.2 when N is sufficiently large and $p^a \leq N/\log^6 N$ we have $f(p^a, N) < 4/\log p^a$. Then for N sufficiently large we have

$$C(N) = \sum_{\sqrt{N}<p^a<N/\log^6 N} \frac{f(p^a, N/p^a)}{p^a} < 4 \sum_{\sqrt{N}<p^a<N/\log^6 N} \frac{1}{p^a \log p^a} = O\left(\frac{1}{\log N}\right)$$

by the Prime Number Theorem.

Again by the Corollary to Proposition 4.2 we have when N is sufficiently large and $p^a \leq N/\log^{3+\epsilon}(N)$ that $f(p^a, N) < 4/\log^{\epsilon/3} p^a = O(1/\log^{\epsilon/3} N)$. Thus

$$\begin{aligned} D(N) &= \sum_{N/\log^6(N)<p^a\leq N/\log^{3+\epsilon}(N)} \frac{f(p^a, N/p^a)}{p^a} \ll \frac{4}{\log^{\epsilon/3} N} \sum_{n/\log^6 N<p^a\leq N/\log^{3+\epsilon} N} \frac{1}{p^a} \\ &= O\left(\frac{\log\log N}{\log^{1+\epsilon/3} N}\right). \end{aligned}$$

4.6 Proof of the Main Theorem

From Propositions 4.1 and 4.3 we have that there is an integer $T(N) \in S_N$ with

$$T(N) > \sum_{1\leq n\leq N} \frac{1}{n} - 1 - G(N),$$

where

$$G(N) = \frac{9}{2} \frac{(\log\log N)^2(1 + o(1))}{\log N}.$$

Thus

$$\left[\sum_{1\le n\le N}\frac{1}{n}-G(N)\right]\le T(N)\le\left[\sum_{1\le n\le N}\frac{1}{n}\right].$$

Therefore there exists an N_0 so that when $N>N_0$ we have

$$0\le|T(N+1)-T(N)|\le\left[\sum_{1\le n\le N+1}\frac{1}{n}\right]-\left[\sum_{1\le n\le N}\frac{1}{n}-G(N)\right]\le 1.$$

It follows that when $N>N_0$ is sufficiently large so that $T(N_0)\le n\le T(N)$, then $n=T(t)$ for some $N_0\le t\le N$. In particular this says that if $T(N_0)\le n\le T(N)$ then there exist integers $1\le x_1<x_2<\cdots<x_k\le N$, for some k, so that

$$n=\frac{1}{x_1}+\cdots+\frac{1}{x_k}.$$

As a consequence of the main result in [35] we have that for N sufficiently large and $1\le n\le T(N_0)$, there exist integers $1\le x_1<x_2<\cdots<x_k\le N_0$, for some k, so that

$$n=\frac{1}{x_1}+\cdots+\frac{1}{x_k}.$$

We conclude that when $N>N_0$,

$$s_N\ge T(N)\ge\left[\sum_{1\le n\le N}\frac{1}{n}-\frac{9}{2}\frac{(\log\log N)^2(1+o(1))}{\log N}\right]$$

as claimed.

Suppose now that $1\le x_1<x_2<\cdots<x_k\le N$ has the property that

$$\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_k}$$

is an integer. We claim that none of the x_i's has a prime factor greater than $N\log\log N/\log N$: for suppose p is such a prime and let $x_{i_1}=pm_1$, $x_{i_2}=pm_2,\ldots,$ $x_{i_l}=pm_l$ be all the x_i's divisible by p. Since $(N\log\log N/\log N)m_i<pm_i\le N$ we have that such $m_i<\log N/\log\log N$ and therefore $l<\log N/\log\log N$. Also, since

$$\frac{1}{x_1}+\cdots+\frac{1}{x_k}$$

is an integer, we must have $p \nmid v$ and yet

$$p \mid m_1 m_2 \cdots m_l \qquad \left(\frac{1}{m_1}+\frac{1}{m_2}+\cdots+\frac{1}{m_l}\right) < l\left(\frac{\log N}{\log\log N}\right)^{l-1} < \left(\frac{\log N}{\log\log N}\right)^{l}$$
$$< \left(\frac{\log N}{\log\log N}\right)^{\frac{\log N}{\log\log N}} < \frac{N\log\log N}{\log N} < p.$$

We conclude that no x_i is divisible by $p > \frac{N\log\log N}{\log N}$. Using this fact and Lemma 4.4 we get the bound

$$\sum_{1\le n\le N}\frac{1}{n} - \sum_{1\le i\le k}\frac{1}{x_i} > \sum_{\substack{\frac{N\log\log N}{\log N}<p\le N\\ mp\le N}} \frac{1}{mp} = \frac{1}{2}\frac{(\log\log N)^2(1+o(1))}{\log N},$$

and so

$$s_N \le \left[\sum_{1\le n\le N}\frac{1}{n} - \frac{1}{2}\frac{(\log\log N)^2(1+o(1))}{\log N}\right]$$

as claimed.

4.7 Proof of the Corollary to the Main Theorem

Fix an $\epsilon > 0$. Let m be a sufficiently large positive integer and select N for which

$$\begin{aligned} m &< \sum_{1\le n\le N}\frac{1}{n} - \left(\frac{9}{2}+\epsilon\right)\frac{(\log\log N)^2}{\log N} \\ &= \log N + \gamma + O\left(\frac{1}{N}\right) - \left(\frac{9}{2}+\epsilon\right)\frac{(\log\log N)^2}{\log N}, \end{aligned}$$

From the Main Theorem we know that for m sufficiently large there exist integers $1 \le x_1 < x_2 < \cdots < x_k \le N$, for some k, so that

$$m = \frac{1}{x_1}+\cdots+\frac{1}{x_k}. \tag{4.5}$$

Since $N > e^{n+O(1)}$, we have that $n < \log N + \gamma - (\frac{9}{2}+\epsilon+o(1))(\log^2 n)/n$. Thus, as long as N satisfies

$$N > e^{n-\gamma+(\frac{9}{2}+\epsilon+o(1))(\log^2 n)/n} = e^{n-\gamma}\left\{1+\left(\frac{9}{2}+\epsilon+o(1)\right)\frac{\log^2 n}{n}\right\},$$

equation (4.5) above has a solution with $x_k < N$. Since $\epsilon > 0$ was arbitrary, the Corollary follows.

Chapter 5

Unit Fractions with Denominators in Short Intervals

5.1 Introduction

Let X_k denote the set

$$\left\{ \{x_1, ..., x_k\} \ : \ \sum_{j=1}^{k} \frac{1}{x_j} = 1, \ 0 < x_1 < \cdots < x_k \right\}.$$

Erdős and Graham (see [12] and [15]) asked the following questions:

1. Is it true that

$$\max\{x_1 \ : \ \{x_1, ..., x_k\} \in X_k\} \ \sim \ \frac{k}{e-1}?$$

Trivially, we have that it is less than or equal to $(1 + o(1))k/(e - 1)$, so all one needs to show is a lower bound of size $(1 + o(1))k/(e - 1)$.

2. Is it true that

$$\min\{x_k - x_1 \ : \ \{x_1, ..., x_k\} \in X_k\} \ \sim \ k?$$

(Note: these two questions were misstated in [12], but are correct in [15].)

In this chapter we will prove the following theorem, which solves these questions of Erdős and Graham for infinitely many k.

Main Theorem *Suppose that $r > 0$ is any given rational number. Then, for all $N > 1$, there exist integers $x_1, ..., x_k$, with*

$$N < x_1 < x_2 < \cdots < x_k \leq \left(e^r + O_r\left(\frac{\log\log N}{\log N}\right)\right) N$$

such that

$$r = \frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_k}.$$

Moreover, the error term $O_r(\log\log N/\log N)$ *is best possible.*

We will now discuss the idea of the proof of the Main Theorem. Let $c > 1$ be the smallest real number where

$$r \le \sum_{N<n<cN} \frac{1}{n} \le r + \frac{1}{cN}.$$

Using the fact that $\sum_{1\le n\le t} 1/n = \log t + \gamma + O(1/t)$ one can show that $c = e^r + O_r(1/N)$. Now suppose

$$\frac{u}{v} = \sum_{N<n<cN} \frac{1}{n}, \quad \text{where } \gcd(u,v) = 1. \tag{5.1}$$

If we had that $u/v = r$, then we would have proved our theorem for this instance of r and N, because $c = e^r + O_r(1/N)$ is well within the error of $O_r(\log\log N/\log N)$ claimed by our theorem. Unfortunately, for large N it will not be the case that $u/v = r$.

To prove the theorem, we first will use a proposition which says that we can remove terms from the sum in (5.1), call them $1/n_1, 1/n_2, ..., 1/n_k$, so that if

$$\frac{u'}{v'} = \frac{u}{v} - \left\{\frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_k}\right\} = \sum_{\substack{N<n<cN \\ n\ne n_1,n_2,\ldots,n_k}} \frac{1}{n}, \text{ where } \gcd(u',v') = 1,$$

then all the prime power factors of v' are $\le N^{1/5}$; moreover, we will have

$$\frac{\log\log N}{\log N} \ll_r \frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_k} \ll_r \frac{\log\log N}{\log N}.$$

We will then couple this with another proposition which says that if s is some rational number whose denominator has all its prime power factors $\le M^{1/4-\epsilon}$, and if $s \gg \log\log\log M/\log M$, then there are integers $M < m_1 < m_2 < \cdots < m_l < e^{(v(\epsilon)+o(1))s}M$, where $v(\epsilon)$ is some constant depending on ϵ, such that

$$s = \frac{1}{m_1} + \frac{1}{m_2} + \cdots + \frac{1}{m_l}.$$

The way we use this second proposition is we let $M = cN$ and

$$s = r - \frac{u'}{v'},$$

so that

$$\frac{\log\log M}{\log M} \ll_r \ s \ \ll_r \frac{\log\log M}{\log M}.$$

Now, all the prime power factors of the denominator of s will be $\leq N^{1/5}$ (when N is sufficiently large). Thus, the hypotheses of this second proposition are met, and so there exist $m_1, ..., m_l$ such that

$$r = s + \frac{u'}{v'} = \sum_{\substack{N<n<cN \\ n\neq n_1,n_2,\ldots,n_k}} \frac{1}{n} \ + \ \sum_{i=1}^{l} \frac{1}{m_i}.$$

All the denominators of these unit fractions will be no larger than

$$e^{(v(\epsilon)+o(1))s} M = e^{(v(\epsilon)+o(1))s+r} N = \left(e^r + O_r\left(\frac{\log\log N}{\log N}\right)\right) N,$$

and, no smaller than N.

The way we will prove that the error term $O_r\left(\log\log N / \log N\right)$ is best possible is by showing that if

$$r = \frac{1}{x_1} + \cdots + \frac{1}{x_k},\ 2 \leq x_1 < \cdots < x_k \ \text{ are integers},$$

then none of the x_i's can be divisible by a prime $p > x_k / \log x_k$ (this idea appears in [12], [19], and [4]). It will turn out that this forces

$$\frac{x_k}{x_1} > e^r \left(1 + \frac{(r+o(1))\log\log x_k}{\log x_k}\right),$$

thus finishing the proof of the Main Theorem.

5.2 Smooth Numbers

In order to even state, let alone prove, the propositions and lemmas needed to prove the Main Theorem, we will need to introduce some notation and definitions concerning smooth numbers. We say that a number n is y-smooth if all of its prime factors are less than or equal to y, and we define the usual smooth number counting function as follows.

$$\psi(N,y) := \#\{n \le N \;:\; n \text{ is } y-\text{smooth}\} = \#\{n \le N \;:\; p|n,\ p \text{ prime } \Rightarrow p \le y\}.$$

Define

$$S(N,y) \;:=\; \{n \le N \;:\; p^a|n,\ p \text{ prime} \Rightarrow p^a \le y\},$$

and let

$$\psi'(N,y) = |S(N,y)|,$$

the number of elements in $S(N,y)$.

In later sections we will need various estimates concerning the $\psi'(N,y)$ and $\psi(N,y)$ functions, and we will use the following lemma to obtain them.

Lemma 5.1 (N.G. de Bruijn) *For any fixed* $\epsilon < 3/5$, *uniformly in the range*

$$y \ge 2, \quad 1 \le u \le \exp\{(\log y)^{3/5-\epsilon}\},$$

we have

$$\psi(N,y) = N\rho(u)\left\{1 + O\left(\frac{\log(u+1)}{\log y}\right)\right\},$$

where $u = \log N/\log y$ *and* $\rho(u)$ *is the unique continuous solution to the differential-difference equation*

$$\begin{cases} \rho(u) = 1, & \text{if } 0 \le u \le 1; \\ u\rho'(u) = -\rho(u-1), & \text{if } u > 1. \end{cases} \tag{5.2}$$

(For a proof of this lemma, see [9].) We can deduce the same estimate for the function $\psi'(N,y)$ by using the following lemma.

Lemma 5.2

$$\sum_{\substack{mp^a \le L \\ p^a \ge y,\ a \ge 2 \\ p \text{ prime}}} \frac{1}{mp^a} = O\left(\frac{\log L}{\sqrt{y}}\right).$$

Proof.

$$\sum_{\substack{mp^a \le L \\ p^a \ge y,\ a \ge 2 \\ p \text{ prime}}} \frac{1}{mp^a} < \sum_{n \ge \sqrt{y}} \sum_{j=2}^{\infty} \sum_{m \le L} \frac{1}{n^j m} \ll \log L \sum_{n \ge \sqrt{y}} \sum_{j=2}^{\infty} \frac{1}{n^j}$$

$$\ll \log L \sum_{n \ge \sqrt{y}} \frac{1}{n^2} \ll \frac{\log L}{\sqrt{y}}.$$

From these last two lemmas we deduce that

$$\psi'(N,y) = \psi(N,y) - O\left(N \sum_{\substack{mp^a \le N \\ p^a \ge y,\ a \ge 2 \\ p \text{ prime}}} \frac{1}{mp^a} \right) = \psi(N,y) - O\left(\frac{N \log N}{\sqrt{y}}\right). \qquad (5.3)$$

Combining this with the previous two lemmas, we have the following final result of this section.

Lemma 5.3 *If* $c \ll 1$ *and* $N \gg_{c,u} 1$*, then*

$$\sum_{\substack{N < n < cN \\ n \in S(N, N^{1/u})}} \frac{1}{n} = \rho(u) \log c + O\left(\frac{1}{\log N}\right).$$

Proof. From Lemma 5.1, Lemma 5.2, and (5.3) we have the following chain of equalities.

$$\begin{aligned}
\sum_{\substack{N < n < cN \\ n \in S(N, N^{1/u})}} \frac{1}{n} &= \left\{ \sum_{\substack{N < n < cN \\ n \text{ is } N^{1/u}\text{-smooth}}} \frac{1}{n} \right\} - O\left(\frac{\psi(N, N^{1/u}) - \psi'(N, N^{1/u})}{N} \right) \\
&= \left\{ \sum_{\substack{N < n < cN \\ n \text{ is } n^{1/u}\text{-smooth}}} \frac{1}{n} - O\left(\sum_{\substack{N^{1/u} < p < (cN)^{1/u} \\ p \text{ prime}}} \sum_{N/p < m < cN/p} \frac{1}{mp} \right) \right\} \\
&\qquad - O\left(\frac{\log(cN)}{N^{1/(2u)}}\right) \\
&= \left\{ \rho(u) \log c + O\left(\frac{1}{\log N}\right) - O\left(\frac{\pi((cN)^{1/u})}{N^{1/u}}\right) \right\} - O\left(\frac{\log(cN)}{N^{1/(2u)}}\right) \\
&= \rho(u) \log c + O\left(\frac{1}{\log N}\right).
\end{aligned}$$

5.3 Proof of the Main Theorem

To prove the Main Theorem we will require the following two propositions, which are the same as those mentioned in the introduction.

Proposition 5.1 *Let $c > 1$. Then, for all N sufficiently large, there exist integers $d_1, ..., d_l$ with $N < d_1 < d_2 < \cdots < d_l < cN$, such that if*

$$\frac{f}{g} = \sum_{\substack{N<n<cN \\ n \neq d_1,...,d_l}} \frac{1}{n}, \tag{5.4}$$

then all the prime power factors of g are $\leq N^{1/5}$, and

$$\frac{\log\log N}{\log N} \ll_c \frac{1}{d_1} + \frac{1}{d_2} + \cdots + \frac{1}{d_l} \ll_c \frac{\log\log N}{\log N}. \tag{5.5}$$

Proposition 5.2 *Suppose $0 < \epsilon < 1/8$ and A and B are positive integers, where $gcd(A, B) = 1$, all the prime power divisors of B are $\leq M^{1/4-\epsilon}$, and $\log\log\log M/\log M \ll A/B \leq 1$. Select $c(M) > 0$ so that*

$$2\frac{A}{B} \leq \sum_{\substack{M \leq n \leq c(M)M \\ n \in S(c(M)M, M^{1/4-\epsilon})}} \frac{1}{n} < 2\frac{A}{B} + \frac{1}{c(M)M}.$$

Then, for all M sufficiently large, there exist integers $n_1, ..., n_k$ with $M \leq n_1 < n_2 < \cdots < n_k \leq c(M)M$, each $n_i \in S(c(M)M, M^{1/4-\epsilon})$, and

$$\frac{A}{B} = \frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_k}.$$

Remark: From Lemma 5.3 we deduce that $c(M) < e^{v(\epsilon)A/B}$, where $v(\epsilon)$ is some function depending only on ϵ. By using a "short interval" version of Lemma 5.1 one can prove an even stronger version of Lemma 5.3, and possibly a stronger version of Proposition 5.2, which would work for all A/B with $1/\log^{1+\epsilon} N \ll A/B < 1$, for any $\epsilon > 0$.

Using these propositions we will now prove the Main Theorem. Let M be the least integer where

$$r \leq \sum_{N<n<M} \frac{1}{n} \leq r + \frac{1}{M}. \tag{5.6}$$

Using the fact that $\sum_{1\leq n\leq x} 1/n = \log x + \gamma + O(1/x)$, it is easy to see that $M/N = e^{r+O(1/N)}$.

Now, from Proposition 5.1, we have that for N sufficiently large, there exist integers $d_1, ..., d_l$ with

$$N < d_1 < d_2 < \cdots < d_l < M = e^{r+O(1/N)}N,$$

such that if

$$\frac{u}{v} := \sum_{\substack{N<n<M \\ n\neq d_1,\ldots,d_l}} \frac{1}{n}, \ \gcd(u,v) = 1,$$

then all the prime power factors of v are $\leq N^{1/5}$. Also, from (5.5) and (5.6) we have that

$$\frac{A}{B} := r - \frac{u}{v} \gg \frac{\log\log N}{\log N}, \gcd(A,B) = 1.$$

We observe that once N is large enough, all the prime power factors of B will be $\leq N^{1/5}$. We conclude from Proposition 5.2 with $\epsilon = 1/20$ that there exist integers $n_1, ..., n_k$ with

$$M \leq n_1 < \cdots < n_k < e^{v(1/20)A/B}M,$$

where $v(1/20)$ is some constant, and such that

$$\frac{A}{B} = \frac{1}{n_1} + \cdots + \frac{1}{n_k},$$

Thus, we have the following representation for r:

$$r = \frac{u}{v} + \frac{A}{B} = \left(\sum_{\substack{N<n<M \\ n\neq d_1,\ldots,d_l}} \frac{1}{n} \right) + \frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_k},$$

where

$$n_k < e^{v(1/20)A/B}M = \left\{1 + O\left(\frac{\log\log M}{\log M}\right)\right\} M = \left\{e^r + O_r\left(\frac{\log\log N}{\log N}\right)\right\} N.$$

This proves the first part of the Main Theorem.

To see that the $O_r(\log\log N/\log N)$ error term is best possible, suppose that

$$r = \frac{U}{V} = \frac{1}{x_1} + \cdots + \frac{1}{x_k}, \text{ where } \gcd(U, V) = 1.$$

We claim that the largest prime p dividing the x_i's satisfies $p < x_k(1 + o(1))/\log x_k$. To see this, fix a prime p and let

$$x_1 = pm_1 < x_2 = pm_2 < \cdots < x_l = pm_l$$

be all the x_i's divisible by p. We have two cases to consider: case 1 is if $p|V$, and case 2 is when $p \nmid V$.

If we are in case 1, where $p|V$, then certainly $p \leq V$, and so $p < x_k(1 + o(1))/\log x_k$, for k sufficienly large (or N sufficiently large). If we are in case 2, where $p \nmid V$, then we must have that $p \nmid B'$ either, where B' is given by

$$\frac{A'}{B'} = \frac{1}{x_1} + \cdots + \frac{1}{x_l} = \frac{1}{p}\left(\frac{1}{m_1} + \cdots + \frac{1}{m_l}\right), \quad \gcd(A', B') = 1.$$

Thus, p divides

$$\begin{aligned} \mathrm{lcm}\{m_1, ..., m_l\}\left\{\frac{1}{m_1} + \cdots + \frac{1}{m_l}\right\} &\leq \mathrm{lcm}\{2, 3, ..., m_l\}\left\{1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{m_l}\right\} \\ &= e^{m_l(1+o(1))}, \end{aligned}$$

and so $p < e^{m_l(1+o(1))}$. From this we deduce that

$$x_k \geq pm_l > p\log p(1 + o(1));$$

or in other words,

$$p < \frac{x_k}{\log x_k}(1 + o(1)).$$

Making use of this bound on p we have that

$$\begin{aligned} r &= \sum_{j=1}^{k} \frac{1}{x_j} \le \sum_{\substack{N<n<cN \\ p|n \Rightarrow p<cN(1+o(1))/\log(cN)}} \frac{1}{n} \\ &= \left(\sum_{N<n<cN} \frac{1}{n} \right) - \left(\sum_{\substack{N<mp<cN \\ p>cN(1+o(1))/\log(cN)}} \frac{1}{mp} \right) \end{aligned} \tag{5.7}$$

The first of this last pair of sums can be estimated using the well-known result $\sum_{x\le n\le y} 1/n = \log x/y + O(1/x)$, which gives

$$\sum_{N<n<cN} \frac{1}{n} = \log c + O\left(\frac{1}{N}\right). \tag{5.8}$$

To estimate the second of the last pair of sums in (5.7), we will need the following lemma, which is proved at the end of this section.

Lemma 5.4 *For $c > 1$ and $\alpha > 0$ we have*

$$\begin{aligned} \sum_{\substack{N<mp<cN \\ p>N/\log^{\alpha} N,\ p \text{ prime}}} \frac{1}{mp} &= \sum_{\substack{N<mp^a<cN \\ p^a>N/\log^{\alpha} N,\ p \text{ prime}}} \frac{1}{mp^a} + O_c\left(\frac{1}{\log N}\right) \\ &= \frac{\alpha(\log c)(\log\log N)}{\log N} + O_c\left(\frac{1}{\log N}\right). \end{aligned}$$

Combining this lemma with (5.7) and (5.8), we have

$$r \le \log c - (\log c + o(1))\frac{\log\log N}{\log N}.$$

Solving for c we find that

$$c \ge e^r\left(1 + \frac{(r+o(1))\log\log N}{\log N}\right).$$

We will now prove Lemma 5.4.

Proof. Using the fact that $\sum_{1\le j\le n} 1/j = \log n + \gamma + O(1/n)$, together with the estimate

$$\sum_{\substack{p\le n \\ p \text{ prime}}} \frac{1}{p} = \log\log n + \kappa + o(1/\log n),$$

where κ is some constant, we have the following chain of inequalities:

$$\begin{aligned}
\sum_{\substack{N<mp\leq cN \\ p>N/\log^\alpha N,\ p \text{ prime}}} \frac{1}{mp} &= \sum_{N/\log^\alpha N<p\leq cN} \frac{1}{p} \sum_{N/p<m\leq cN/p} \frac{1}{m} \\
&= \sum_{N/\log^\alpha N<p\leq cN} \frac{1}{p}\left\{\log\left(\frac{cN}{p}\right) - \log\left(\frac{N}{p}\right) + O\left(\frac{p}{N}\right)\right\} \\
&= \sum_{N/\log^\alpha N<p\leq cN} \frac{1}{p}\left\{\log c + O\left(\frac{p}{N}\right)\right\} \\
&= \log c \sum_{N/\log^\alpha N<p\leq cN} \frac{1}{p} + O\left(\frac{\pi(cN)}{N}\right) \\
&= \log c\left\{\log\log cN - \log\log\left(\frac{N}{\log^\alpha N}\right) + o(1/\log N)\right\} \\
&\qquad +O_c\left(\frac{1}{\log N}\right) \\
&= \frac{\alpha(\log c)(\log\log N)}{\log N} + O_c\left(\frac{1}{\log N}\right),
\end{aligned}$$

as claimed. The proof for the sum over prime power p^a, instead of prime powers p, is almost exactly the same.

5.4 Proof of Proposition 5.1

Let $p_1 < p_2 < \cdots < p_h$ be all the primes in $(N^{1/5}, N/\log^{10} N)$. Define

$$S \ := \ (N,\ cN) \ \cap \ \mathbb{Z},$$

$$\begin{aligned}
S_{h+1} \ := \ S \ \setminus \ \Big(&\{mp \ : \ p \text{ prime}\ ,\ p > N/\log^{10} N\} \\
&\cup \{mp^a \ : \ p \text{ prime}\ ,\ a \geq 2,\ p^a > N^{1/5}\}\Big),
\end{aligned}$$

and let

$$\frac{u_{h+1}}{v_{h+1}} = \sum_{n\in S_{h+1}} \frac{1}{n}, \text{ where } \gcd(u_{h+1}, v_{h+1}) = 1.$$

Notice that v_{h+1} has no prime divisor $\geq N/\log^{10} N$; moreover, v_{h+1} has no prime power factors $\geq N/\log^{10} N$, for N sufficiently large, since the only prime power divisors of elements of S that are $\geq N^{1/5}$ are primes. We also have that

$$\sum_{n \in S \setminus S_{h+1}} \frac{1}{n} = \sum_{\substack{N \leq mp \leq cN \\ p \geq N/\log^{10} N}} \frac{1}{mp} + O\left(\sum_{\substack{mp \leq cN \\ p \geq N^{1/5},\ a \geq 2 \\ p \text{ prime}}} \frac{1}{mp^a} \right). \tag{5.9}$$

The first of these last two sums can be estimated using Lemma 5.4, which gives

$$\sum_{\substack{N \leq mp \leq cN \\ p \geq N/\log^{10} N}} \frac{1}{mp^a} = \frac{(10 \log c + o(1)) \log\log N}{\log N},$$

and the second of the last two sums can be estimated using Lemma 5.2, which gives

$$\sum_{\substack{mp^a \leq cN \\ p \geq N^{1/5},\ a \geq 2 \\ p \text{ prime}}} \frac{1}{mp^a} = O\left(\frac{\log(cN)}{N^{1/10}} \right).$$

Combining the last two displayed equations with (5.9), we deduce that

$$\sum_{n \in S \setminus S_{h+1}} \frac{1}{n} = \frac{(10 \log c + o(1)) \log\log N}{\log N}.$$

Starting with the prime p_h we will successively construct sets

$$S_h \supseteq S_{h-1} \supseteq S_{h-2} \supseteq \cdots \supseteq S_1,$$

where if

$$\frac{u_i}{v_i} = \sum_{n \in S_i} \frac{1}{n},$$

$\gcd(u_i, v_i) = 1$, then all the prime factors of v_i are smaller than p_i, for all $i = 1, 2, .., h$; moreover, we will construct these sets in such a way that

$$\sum_{n \in S_{i+1} \setminus S_i} \frac{1}{n} \ll \frac{1}{p_i \log N}, \text{ for } i = 1, 2, ..., h.$$

If we can accomplish this, then if we let $\{d_1, ..., d_l\} = S \setminus S_1$, we will have

$$\begin{aligned} \frac{1}{d_1} + \frac{1}{d_2} + \cdots + \frac{1}{d_l} &= \sum_{n \in S \setminus S_{h+1}} \frac{1}{n} + \sum_{j=2}^{h+1} \sum_{n \in S_j \setminus S_{j-1}} \frac{1}{n} \\ &= \sum_{n \in S \setminus S_{h+1}} \frac{1}{n} + O\left(\sum_{j=1}^{h} \frac{1}{p_j \log N} \right) \\ &\ll \frac{\log\log N}{\log N}, \end{aligned}$$

and

$$\frac{1}{d_1} + \frac{1}{d_2} + \cdots + \frac{1}{d_l} \geq \sum_{n \in S \setminus S_{h+1}} \frac{1}{n} = \frac{(10 \log c + o(1)) \log \log N}{\log N}.$$

Thus, (5.5) will be satisfied. We will also have that

$$\sum_{\substack{n \in S \\ n \neq d_1, \ldots, d_l}} \frac{1}{n} = \frac{u_1}{v_1},$$

where all of the prime factors of v_1 are smaller than $N^{1/5}$; moreover, all the prime power factors of v_1 will be smaller than $N^{1/5}$, since the only prime powers $\geq N^{1/5}$ that can divide elements of S are primes. Thus, (5.4) will be satisfied, and so if we can construct these sets S_i, Proposition 5.1 will be proved.

Suppose, for proof by induction, we have constructed the sets S_j where $2 \leq i \leq j \leq h+1$. Then, all the prime factors of v_i are $\leq p_{i-1}$. If $p_{i-1} \nmid v_i$, we just let $S_{i-1} := S_i$, and then all the prime factors of v_{i-1} are smaller than p_{i-1}.

If $p_{i-1} | v_i$, then $p_{i-1} || v_i$, since the only prime power factors of elements of S that are $\geq N^{1/5}$ are primes. We will use Proposition 5.2 to construct S_{i-1} as follows: Using Bertrand's Postulate, let q be the smallest prime in $[\log N, 2 \log N]$, and set $M = N/(q p_{i-1}) > (\log^9 N)/2$. Let

$$B = \mathrm{lcm}\{n \leq M^{1/5}\} > \mathrm{lcm}\{n \leq (\log N)^{9/5}/2^{1/5}\} > 2 c p_{i-1} M$$

(which will be true for M sufficiently large), and let A be the largest integer $\leq c'B/2$ where

$$c' = \sum_{\substack{M < n < cM \\ n \in S(cM, M^{1/5})}} \frac{1}{n} = \rho(5) \log c + O\left(\frac{1}{\log M}\right),$$

(which follows from Lemma 5.3) and

$$A \equiv q B u_i (v_i / p_{i-1})^{-1} \pmod{p_{i-1}}.$$

Since $B > 2 c p_{i-1} M$, and since $A \in [c'B/2 - p_{i-1}, c'B/2]$, we have that

$$2\frac{A}{B} \leq c' < 2\frac{A}{B} + \frac{2 p_{i-1}}{B} < 2\frac{A}{B} + \frac{1}{cM},$$

for N sufficiently large. From Proposition 5.2, there exists $n_1, ..., n_k$, with $M < n_1 < n_2 < \cdots < n_k < cM$ where each $n_i \in S(cM, M^{1/5})$ and

$$\frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_k} = \frac{A}{B}.$$

Now, we claim that we can let

$$S_{i-1} \;=\; S_i \setminus \{qp_{i-1}n_j,\ 1 \le j \le k\}.$$

Notice that these integers we remove all have their largest prime divisor equal to p_{i-1}, so that we remove different numbers for different sets S_j. If we let S_{i-1} be defined in this way, then, since $p_{i-1}|v_i$,

$$\frac{u_{i-1}}{v_{i-1}} = \sum_{n \in S_{i-1}} \frac{1}{n} = \frac{u_i}{v_i} - \frac{A}{qp_{i-1}B} = \frac{qBu_i - Av_i/p_{i-1}}{v_i qB}.$$

So, $u_{i-1}|(qBu_i - Av_i/p_{i-1})$ and $v_{i-1}|v_i qB$. Since $qBu_i - Av_i/p_{i-1} \equiv 0 \pmod{p_{i-1}}$, we must have that $v_{i-1}|v_i qB/p_{i-1}$. Now v_i/p_{i-1} cannot be divisible by p_{i-1}, since v_i is not divisible by the square of a prime $> N^{1/5}$, and so v_i/p_{i-1} has all its prime power divisors $< p_{i-1}$; also, B is not divisible by p_{i-1}, since $B = \mathrm{lcm}\{2, 3, ..., M^{1/5}\}$, and $M^{1/5}$ is less than $N^{1/5} < p_{i-1}$. So, all the prime divisors of v_{i-1} are $< p_{i-1}$. We also have that

$$\sum_{n \in S_i \setminus S_{i-1}} \frac{1}{n} < \sum_{M < n < cM} \frac{1}{qp_{i-1}n} \ll \frac{\log c}{p_{i-1}q} \ll_c \frac{1}{p_{i-1}\log N},$$

and so S_{i-1} satisfies all the requisite properties. We conclude that all the sets S_j, $j = 1, ..., h+1$, can be constructed, and so Proposition 5.1 follows.

5.5 Proof of Proposition 5.2

Let

$$P \;:=\; \mathrm{lcm}\{2, 3, ..., [M^{1/4-\epsilon}]\} = e^{M^{1/4-\epsilon}(1+o(1))},$$

where this last equality follows from the Prime Number Theorem. Let $M \le m_1 < m_2 < \cdots < m_l \le c(M)M$ be all the divisors of P lying in $[M, c(M)M]$; that is, all the integers in $S(c(M)M, M^{1/4-\epsilon})$ in the interval $[M, c(M)M]$. If $Y|P$, we have the following identity.

$$\frac{1}{P}\sum_{h=-P/2}^{P/2-1} e\left(\frac{Xh}{Y}\right) = \begin{cases} 1, & \text{if } Y|X \\ 0, & \text{if } Y \nmid X. \end{cases}$$

Thus, if $B|P$, one can deduce that

$$\begin{aligned}\#\Big\{\{n_1, ..., n_k\} \quad &\subseteq \{m_1, ..., m_l\}, k \text{ variable } : 1/n_1 + \cdots + 1/n_k = A/B\Big\} \\ &\ge \frac{1}{P}\sum_{h=-P/2}^{P/2-1} e\left(\frac{-Ah}{B}\right)\prod_{j=1}^{l}\left\{1 + e\left(\frac{h}{m_j}\right)\right\} - 2,\end{aligned}$$

(The reason for subtracting 2 in the above equation is that when $A/B = 1$, the exponential sum also counts the extraneous representations $1/n_1 + \cdots + 1/n_k = 0$ and 2.)

Let

$$\begin{aligned} F(h) &:= \prod_{j=1}^{l}\left\{1 + e\left(\frac{h}{m_j}\right)\right\} \\ &= e\left(\frac{h}{2}\left\{\frac{1}{m_1} + \cdots + \frac{1}{m_l}\right\}\right)\left(2^l \prod_{j=1}^{l}\cos(\pi h/m_j)\right). \end{aligned} \tag{5.10}$$

Upon substituting this into our equation above this gives

$$\begin{aligned}\#\Big\{\{n_1, ..., n_k\} \quad &\subseteq \{m_1, ..., m_l\}, k \text{ variable } : 1/n_1 + \cdots + 1/n_k = A/B\Big\} \\ &\ge \frac{1}{P}\left(\sum_{h=-P/2}^{P/2-1} e(-Ah/B)F(h)\right) - 2. \end{aligned} \tag{5.11}$$

We will now try to find a lower bound for (5.11). To do this we will show that

$$|F(h)| < \frac{2^l}{2P}, \quad \text{for } M/2 < |h| \le P/2, \tag{5.12}$$

and that

$$\sum_{1\le h\le M/2} e(-Ah/B)F(h) + e(Ah/B)F(-h) > 0, \tag{5.13}$$

from which we deduce

$$\sum_{0\le |h|\le M/2} e(-Ah/B)F(h) > F(0) = 2^l.$$

From this, (5.11), and (5.12), it then follows that

$$\begin{aligned}\#\Big\{\{n_1,...,n_k\} \quad &\subseteq \{m_1,...,m_l\},\ k \text{ variable } :\ 1/n_1+\cdots+1/n_k = A/B\Big\} \\ > \quad & \frac{2^{l-1}}{P} - 2 = 2^{l-O(M^{1/4-\epsilon})},\end{aligned}$$

which is exponential in l since

$$l \gg_\epsilon M\frac{A}{B} \gg \frac{M\log\log\log M}{\log M}.$$

To establish (5.13), we first observe from (5.10) that

$$\begin{aligned}\mathrm{Arg}\{e(-Ah/B)F(h)\} = \frac{-2\pi Ah}{B} \quad &+ \quad \pi h\left\{\frac{1}{m_1}+\cdots+\frac{1}{m_l}\right\} \\ &+ \quad \mathrm{Arg}\left\{\prod_{j=1}^{l}\cos(\pi h/m_j)\right\}.\end{aligned} \tag{5.14}$$

Using the fact that

$$\frac{1}{m_1}+\cdots+\frac{1}{m_l} = 2\frac{A}{B}+\delta,$$

where

$$0\le\delta\le\frac{1}{c(M)M},$$

together with the fact that each m_j is $\ge M$, we have

$$\left|\frac{-2\pi Ah}{B}+\pi h\left\{\frac{1}{m_1}+\cdots+\frac{1}{m_l}\right\}\right| = \pi\delta|h| < \frac{\pi|h|}{M}\le\frac{\pi}{2}, \tag{5.15}$$

whenever

$$|h|\le\frac{M}{2}.$$

Also, for such h, we observe that

$$\cos(\pi h/m_j) \geq \cos(\pi/2) = 0, \text{ for } j = 1, 2, ..., l,$$

since $m_j \geq M$ for all j. Hence,

$$\text{Arg}\left\{\prod_{j=1}^{l} \cos(\pi h/m_j)\right\} = 0.$$

Using this, together with (5.14) and (5.15), we find that

$$|\text{Arg}\{e(-Ah/B)F(h)\}| < \frac{\pi}{2}, \text{ whenever } |h| < \frac{M}{2}.$$

Thus, for such h we have

$$e(-Ah/B)F(h) + e(Ah/B)F(-h) > 0,$$

and so (5.13) follows.

In order to establish (5.12), we will need the following lemma, which will be proved in the next section of the paper:

Lemma 5.5 *Suppose* $0 < \epsilon < 1/8$. *Let* $m_1 < m_2 < \cdots < m_l$ *be all the integers in* $[M, (1 + 1/\log M)M]$ *where each* $m_i \in S((1 + 1/\log M)M, M^{1/4-\epsilon})$. *Then for* M *sufficiently large and* h *real, either*

1. There are $\gg M^{3/4}$ m_i*'s which do not divide any integer in* $I := (h - M^{3/4}, h + M^{3/4})$, *or*

2. There is an integer in this interval which is divisible by $P := \text{lcm}\{p^a \leq M^{1/4-\epsilon} : p \text{ prime}\}$.

From this lemma, it follows that if

$$\frac{M}{2} \leq |h| \leq P/2,$$

and if

$$Z(c_1) \;=\; \#\left\{m_j,\; j=1,...,l \;:\; ||h/m_j|| > \frac{c_1}{M^{1/4}}\right\},$$

where $||t||$ denote the distance to the nearest integer from t, then for some constants $c_1, c_2 > 0$ we will have for all M sufficiently large,

$$Z(c_1) > c_2 M^{3/4}.$$

For these integers m_j counted by $Z(c_1)$, we will have that

$$\begin{aligned} |\cos(\pi h/m_j)| &= |\cos(||\pi h/m_j||)| \\ &< \left|\cos\left(\pi c_1/M^{1/4}\right)\right| \\ &= 1 - \frac{1}{2}\frac{\pi^2 c_1^2}{M^{1/2}} + O\left(\frac{1}{M}\right). \end{aligned}$$

From this and (5.10) it follows that for such h

$$|F(h)| < 2^l\left(1 - \frac{1}{2}\frac{\pi^2 c_1^2}{M^{1/2}} + O\left(\frac{1}{M}\right)\right)^{Z(c_1)} \ll 2^l e^{-\pi^2 c_2 c_1^2 M^{1/4}/2} = o\left(\frac{2^l}{P}\right).$$

This establishes (5.12) and thus proves the proposition.

5.6 Proof of Lemma 5.5

For each integer n satisfying

$$M^{3/4}\log^3 M < n < 2M^{3/4}\log^3 M, \quad \text{and} \quad n \in S(2M^{3/4}\log^3 M, M^{1/4-\epsilon}), \qquad (5.16)$$

define

$$M(n) \;:=\; \{m_j \;:\; m_j = nq,\; \text{where } \omega(q) \le 3.\}$$

We claim that lcm $M(n) = P$ for all such n. We will show below that the truth of this claim implies that either:

A. There is an n satisfying (5.16) such that every integer of $M(n)$ divides a single integer in I, which together with the assumption lcm $M(n) = P$, gives us case 2 in the claim of our lemma, or

B. For each n satisfying (5.16), there is an integer $m_{\alpha(n)} \in M(n)$ which does not divide any integer in $(h - M^{3/4}, h + M^{3/4})$.We will assume that case B is true and show that it implies case 1 in the claim of our lemma (and thus if we can show that lcm $M(n) = P$ and that either A or B is true, we may conclude that either case 1 or case 2 in our lemma is true):

The first thing to notice is that from (5.3) we know there are at least $c_\epsilon M^{3/4}\log^3 M$ integers n satisfying (5.16). If all of the $m_{\alpha(n)}$'s as indicated in case B were distinct, then we would have that there are at least $c_\epsilon M^{3/4}\log^3 M$ m_j's not dividing any integer in $(h - M^{3/4}, h + M^{3/4})$, which is the first possibility claimed by our lemma; however, it is not necessarily the case that the $m_{\alpha(n)}$'s are distinct. To overcome this difficulty, we will now show that no m_i can lie in too many of the sets $M(n)$: Let

$$\begin{aligned} D(M) &:= \max_{m_i} \#\Big\{ n \ : \ n \text{ satisfies (5.16) and } m_i \in M(n) \Big\} \\ &\leq \max_{m_i} \#\{ q \ : \ q | m_i,\ \omega(q) \leq 3 \} \\ &= o\left(\log^3 M\right). \end{aligned}$$

From this we have that

$$\begin{aligned} \#\Big\{ m_{\alpha(n)} \quad &: \quad n \text{ satisfies (5.16)} \Big\} \\ &\geq \frac{\psi(2M^{3/4}\log^3 M, M^{1/4-\epsilon}) - \psi(M^{3/4}\log^3 M, M^{1/4-\epsilon})}{D(M)} \gg M^{3/4}. \end{aligned}$$

Thus, there are $\gg M^{3/4}$ m_j's which do not divide any integer in $(h - M^{3/4}, h + M^{3/4})$, which covers case 1 claimed by our lemma.

We now will show that if lcm $M(n) = P$ for all n satisfying (5.16), then either case A or case B above must be true. So, let us assume then that lcm $M(n) = P$ for all n satisfying (5.16). If case B is true, then we are done. So, let us assume that case B is false. Then, we must have that there is an n satisfying (5.16) such that each member of $M(n)$ divides an integer in I. Since each such member is divisible by $n \geq M^{3/4}\log^3 M$, which is greater than the length of I, we must have that all such members divide the same integer in I. Thus, case A is true.

To finish the proof of our lemma, we now show that lcm $M(n) = P$ for all n satisfying (5.16). Fix an n satisfying (5.16) and let $p^a \leq M^{1/4-\epsilon}$ be the largest power of the prime p that is $\leq M^{1/4-\epsilon}$. Let p^e be the exact power of p which divides n. Thus, $e \leq a$. We will show there exists an $m_j \in M(n)$ with

$$m_j = np^{a-e}l_1l_2, \quad \text{where } l_1 \text{ and } l_2 \text{ are primes with } \gcd(l_1l_2, n) = 1,$$

which will imply that m_j is divisible by p^a, and thus $p^a|$ lcm $M(n)$. Such an m_j exists if we can just find primes $l_1, l_2 \leq M^{1/4-\epsilon}$ which satisfy

$$\sqrt{\frac{M}{np^{a-e}}} \leq l_1 < l_2 \leq \sqrt{\left(1 + \frac{1}{\log M}\right)\frac{M}{np^{a-e}}}, \quad \gcd(l_1l_2, n) = 1. \tag{5.17}$$

To see that it is possible to find l_1 and l_2 we first observe that the lower limit of the interval in (5.17) is

$$\sqrt{\frac{M}{np^{a-e}}} \gg \sqrt{\frac{M}{(M^{3/4}\log^3 M)M^{1/4-\epsilon}}} = \frac{M^{\epsilon/2}}{\log^{3/2} M},$$

and the length of the interval is the multiple $\sqrt{1 + 1/\log M} - 1 \gg 1/\log M$ of this lower limit. By the Prime Number Theorem, there are $\gg M^{\epsilon/2}/(\epsilon \log^{7/2} M)$ primes in this interval, and so for M sufficiently large there must be two of them $l_1 < l_2$ which do not divide $n < 2M^{3/4}\log^3 M$. These two primes therefore satisfy (5.17). To see that $l_1, l_2 < M^{1/4-\epsilon}$, we observe that the upper limit of the interval in (5.17) satisfies

$$\sqrt{\left(1 + \frac{1}{\log M}\right)\frac{M}{np^{a-e}}} < \sqrt{\frac{2M}{n}} \leq \sqrt{\frac{2M}{M^{3/4}\log^3 M}} = \frac{\sqrt{2}M^{1/8}}{\log^{3/2} M} < M^{1/4-\epsilon},$$

for M sufficiently large and $0 < \epsilon < 1/8$. Thus, we can find l_1 and l_2 as claimed, and so our lemma is proved.

Chapter 6

On a Coloring Conjecture about Unit Fractions

6.1 Introduction

We will prove a result on unit fractions which has the following two corollaries.

Corollary 6.1 *There exists a constant b so that for every partition of the integers in $[2, b^r]$ into r classes, there is always one class containing a subset S with the property $\sum_{n \in S} 1/n = 1$.*

Corollary 6.2 *Let S be a subset of $\{1, 2, ..., x\}$ with $\lambda(x) = \sum_{n \in S} 1/n$ maximal such that there is no subset of S with sum of reciprocals equal to 1. Then $\lambda(x) < c \log x$, for some constant $c < 1$.*

In fact, we will show for Corollary 6.1 that b may be taken to be e^{167000}, if r is sufficiently large, though we believe that b may be taken to be much smaller; also, note that b cannot be taken to be smaller than e, since the integers in $[2, e^{r-o(r)}]$ can be placed into r classes in such a way that the sum of reciprocals in each class is just under 1.

Corollary 6.1 implies the result mentioned in the abstract and so resolves an unsolved problem of Erdős and Graham, which appears in [12], [15], and [22].

We will need to introduce some notation and definitions in order to state the Main Theorem, as well as the Propositions and Lemmas in later sections: For a given set of integers C, let $\mathcal{Q}_C$ denote the set of all the prime power divisors of elements of C, and let $\Sigma(C) = \sum_{q \in \mathcal{Q}_C} 1/q$. Define $\mathcal{C}(X, Y; \theta)$ to be the integers in $[X, Y]$ all of whose prime power divisors are $\leq X^\theta$, and let $\mathcal{C}'(X, Y; \theta)$ be those integers $n \in \mathcal{C}(X, Y; \theta)$ such that $\omega(n) \sim \Omega(n) \sim \log\log n$, where $\omega(n)$ and $\Omega(n)$ denote the number of prime divisors and the number of prime power divisors of n, respectively.

Our Main Theorem, then, is as follows.

Main Theorem *Suppose $C \subseteq \mathcal{C}'(N, N^{1+\delta}; \theta)$, where $\theta, \delta > 0$, and $\delta + \theta < 1/4$. If $N \gg_{\theta,\delta} 1$ and*

$$\sum_{n \in C} \frac{1}{n} > 6,$$

then there exists a subset $S \subset C$ for which $\sum_{n \in S} 1/n = 1$.

To prove the Corollaries, we will show in the next section that for r sufficiently large,

$$\sum_{n \in \mathcal{C}'(N, N^{1+\delta}; 1/4.32)} \frac{1}{n} \;>\; 6r, \tag{6.1}$$

where $N = e^{163550r}$ and $N^{1+\delta} = e^{166562r}$. Thus, if we partition the integers in $[2, e^{167000r}]$ into r classes, then for r sufficiently large, one of the classes C satisfies the hypotheses of the Main Theorem, and so our Corollary 6.1 follows. We note that (6.1) also implies that there exists a constant $c > 0$ so that if S is any subset of $[N, N^{1+\delta}]$ with $\sum_{n \in S} 1/n > cr$, then

$$\sum_{n \in S \cap \mathcal{C}'(N, N^{1+\delta}; 1/4.32)} \frac{1}{n} \gg r,$$

(because $\sum_{N < n < N^{1+\delta}} 1/n \gg r$) and so from our Main Theorem, S constains a subset whose sum of reciprocals equals 1 (when r is sufficiently large), which proves the Corollary 6.2.

The key idea in the proof of the Main Theorem is to construct a subset of C with useable properties. These are summarized in the following Proposition which is proved in section 6.4.

Proposition 6.1 *If $C \subseteq \mathcal{C}'(N, N^{1+\delta}; \theta)$ with $\delta + \theta < 1/4$, then there exists a subset $D \subset C$ such that*

$$\sum_{n \in D} \frac{1}{n} \in [2 - 3/N, 2), \tag{6.2}$$

and which has the following property: If I is an interval of length $N^{3/4}$ for which there are less than $N^{1-\theta}/(\log\log N)^2$ elements of D that do not divide any element of I, then every element of D divides one single element of I.

The sum of the reciprocals of the elements of D is < 2 by (6.2), so if there is a subset S of D for which $\sum_{n \in S} 1/n$ is an integer then that sum equals 1 or S is the empty set. Now if x is an integer and

$$P \ := \ \mathrm{lcm}\{n \in D\},$$

then $(1/P) \sum_{h \pmod P} e(hx/P) = 1$ if x/P is an integer, and is 0 otherwise, where $e(t) = e^{2\pi i t}$. Combining these remarks we deduce that

$$\#\left\{ S \subset D : \sum_{n \in S} 1/n = 1 \right\} = \left(\frac{1}{P} \sum_{-P/2 < h \leq P/2} E(h) \right) - 1, \tag{6.3}$$

where

$$E(h) \ := \ \prod_{n \in D} \left(1 + e(h/n)\right).$$

Now,

$$E(h) \ = \ e\left(\frac{h}{2} \left\{ \sum_{n \in D} \frac{1}{n} \right\} \right) \left(2^{|D|} \prod_{n \in D} \cos(\pi h/n) \right), \tag{6.4}$$

so that

$$\mathrm{Arg}(E(h)) \ = \ \pi h \left\{ \sum_{n \in D} \frac{1}{n} \right\} \in (2\pi h - \pi/2, 2\pi h + \pi/2),$$

if $|h|$ is an integer $< N/6$; and therefore $E(h) + E(-h) > 0$ for this case. Thus we deduce that

$$\sum_{|h| < N/6} E(h) > E(0) = 2^{|D|}.$$

For h in the range $N/6 \le |h| \le P/2$, we will use Proposition 6.1 to show that

$$|E(h)| < \frac{2^{|D|-1}}{P}, \tag{6.5}$$

so that, by the last two displayed equations,

$$\frac{1}{P}\sum_{-P/2<h\le P/2} E(h) > \frac{1}{P}\left(2^{|D|} - \sum_{N/6\le|h|\le P/2}\frac{2^{|D|-1}}{P}\right) > \frac{2^{|D|-1}}{P} > 1,$$

since $|D| \ge \sum_{n\in D} N/n \ge 2N-3$, and since

$$P < \left(N^{\theta}\right)^{\pi(N^{\theta})} \ll e^{(1+o(1))N^{\theta}} = o\left(2^{|D|}\right), \tag{6.6}$$

by the prime number theorem. The Main Theorem then follows.

We will now see how (6.5) follows from Proposition 6.1. If $|h| \in [N/6, P/2]$ then $I := [h - N^{3/4}/2, h + N^{3/4}/2]$ does not contain any integer divisible by every element of D, since $P = \mathrm{lcm}_{n\in D} n$ is bigger than every element in I. Therefore, by Proposition 6.1, there are at least $N^{1-\theta-o(1)}$ elements $n \in D$ which do not divide any integer in I. For such n we will have that $||h/n|| > N^{3/4}/2n > 1/(2N^{1/4+\delta})$ (where $||t||$ denotes the distance from t to the nearest integer to t). Thus,

$$\begin{aligned}\left|\prod_{n\in D}\cos(\pi h/n)\right| &< \left|\cos\left(\frac{\pi}{2N^{1/4+\delta}}\right)\right|^{N^{1-\theta-o(1)}} \\ &< \left(1 - \frac{\pi^2}{8N^{1/2+2\delta}} + O\left(\frac{1}{N}\right)\right)^{N^{1-\theta-o(1)}} \\ &< \exp\left(-(\pi^2/8)N^{1/2-2\delta-\theta-o(1)}\right) < \frac{1}{2P},\end{aligned}$$

by (6.6) since $\delta + \theta < 1/4$, and so (6.5) follows from (6.4).

The rest of the paper is dedicated to proving Proposition 6.1.

6.2 Normal Integers with Small Prime Factors

We will need the following result of Dickman from [1].

Lemma 6.1 *Fix $u_0 > 0$. For any u, $0 < u < u_0$ we have*

$$\#\{n \le x : p|n \Rightarrow p \le x^{1/u}\} \sim x\rho(u),$$

where $\rho(u)$ is the unique, continuous solution to the differential difference equation

$$\begin{cases} \rho(u) = 1, & \text{if } 0 \le u \le 1 \\ u\rho'(u) = -\rho(u-1), & \text{if } u > 1. \end{cases}$$

From this lemma and partial summation we have, for a fixed u and δ,

$$\sum_{\substack{N<n<N^{1+\delta} \\ p^a||n \Rightarrow p^a \le N^{1/u}}} \frac{1}{n} \sim \frac{\log N}{u} \int_u^{u(1+\delta)} \rho(w)dw.$$

Using this, a numerical calculation shows for $N = \exp(163550r)$, $\theta = 1/u = 1/4.32$, and $\delta = 1/4 - \theta - 0.0001$ that

$$\sum_{\substack{N<n<N^{1+\delta} \\ p^a||n \Rightarrow p^a \le N^{\theta}}} \frac{1}{n} > 6.0001r.$$

Combining this with the well-known fact that almost all integers $n \le x$ satisfy $\omega(n) \sim \Omega(n) \sim \log\log n$, so that

$$\sum_{\substack{N<n<N^{1+\delta} \\ \omega(n) \text{ or } \Omega(n) \not\sim \log\log N}} \frac{1}{n} = o(r),$$

we have that (6.1) follows.

6.3 Technical Lemmas and their Proofs

Lemma 6.2 *If w_1 and w_2 are distinct integers which both lie in an interval of length $\le N$, then*

$$\sum_{p^a|\gcd(w_1,w_2)} \frac{1}{p^a} < \sum_{p|\gcd(w_1,w_2)} \frac{1}{p} + O(1) < (1+o(1))\log\log\log N.$$

Proof of Lemma 6.2. Let $G=\gcd(w_1,w_2)$. We have that $G \leq |w_1 - w_2| < N$, since $G||w_1 - w_2|$; also, $\omega(G) = o(\log N)$, since $\omega(n) = o(\log N)$ uniformly for $n \leq N$. Now, by the Prime Number Theorem, we have $\pi(\log N \log\log N) \gg \log N > w(G)$, for N sufficiently large, and so

$$\sum_{\substack{p|G \\ p \text{ prime}}} \frac{1}{p} < \sum_{\substack{p \leq \log N \log\log N \\ p \text{ prime}}} \frac{1}{p} < (1+o(1)) \log\log\log N.$$

Lemma 6.3 *If $H \subseteq \mathcal{C}(N, N^{1+\beta}; 1)$, $\beta > 0$, satisfies $\sum_{n \in H} 1/n > 1/(\log N)^{o(1)}$, and $\omega(n) \sim \log\log n$, for every $n \in H$, then*

$$\Sigma(H) > (e^{-1} - o(1)) \log\log N.$$

Proof of Lemma 6.3. From the hypotheses of the lemma, together with the fact that $t! > (t/e)^t$ for $t \geq 1$, we have that

$$\begin{aligned} \frac{1}{(\log N)^{o(1)}} &< \sum_{n \in H} \frac{1}{n} < \sum_{\substack{n \,:\, p^a | n \Rightarrow p^a \in \mathcal{Q}_H \\ \omega(n) \sim \log\log n \sim \log\log N}} \frac{1}{n} < \sum_{t \sim \log\log N} \frac{\Sigma(H)^t}{t!} \\ &< \sum_{t \sim \log\log N} \left(\frac{\Sigma(H) e}{t} \right)^t = \left(\frac{\Sigma(H)(e + o(1))}{\log\log N} \right)^{(1+o(1))\log\log N}, \end{aligned}$$

and so $\Sigma(H)$ satisfies the conclusion to Lemma 6.3.

6.4 Proof of Proposition 6.1

Before we prove Proposition 6.1, we will need two more Propositions.

Proposition 6.2 *Suppose that $J \subseteq \mathcal{C}(N, \infty; \theta)$, where $\theta < 1$, and $\sum_{n \in J} 1/n \geq \alpha > \nu$. If $N \gg_{\alpha,\nu,\theta} 1$, then there is a subset $E \subset J$ such that*

$$\sum_{n \in E} \frac{1}{n} \in \left[\nu - \frac{1}{N}, \nu \right); \text{ and,} \tag{6.7}$$

$$\sum_{\substack{n \in E \\ q | n}} \frac{1}{n} > \frac{\min\{\nu, \alpha - \nu\}}{5q \log\log N}, \text{ for all } q \in \mathcal{Q}_E, \tag{6.8}$$

Proposition 6.3 *Suppose that $E \subseteq \mathcal{C}'(N, N^{1+\delta}; \theta)$, $0 < \theta < 1/4$, satisfies (6.7) and (6.8). If all but at most $N^{1-\theta}/(\log\log N)^2$ elements of E divide some element of an interval $I := [h - N^{3/4}/2, h + N^{3/4}/2]$, then either*

A. There is a single integer in I divisible by all elements of E, or

B. There exist distinct integers $w_1, w_2 \in I$, such that

$$\#\{n \in E \ : \ n \nmid w_1 \quad \text{and } n \nmid w_2\} < \frac{2N^{1-\theta}}{(\log\log N)^2}, \tag{6.9}$$

$$\operatorname{lcm}\{n \in E\} = \operatorname{lcm}\{q \in \mathcal{Q}_E\} | w_1 w_2, \text{ and} \tag{6.10}$$

$$(e^{-1} - o(1))\log\log N < \sum_{\substack{q|w_i \\ q \in \mathcal{Q}_E}} \frac{1}{q} < (1 - e^{-1} + o(1))\log\log N, \text{ for } i = 1 \text{ and } 2. \tag{6.11}$$

These Propositions will be proved in the next two sections of the paper. To prove Proposition 6.1, we iterate the following procedure:

1. Set $j = 0$ and let $C_0 := C$.

2. Use Proposition 6.2 with $J = C_j$, $\alpha = \sum_{n \in C_j} 1/n > 2$, and $\nu = 2$, to produce a subset E satisfying (6.7) and (6.8).

3. If case A of Proposition 6.3 holds for every real number h satisfying the hypotheses of Propositon 6.3, then we can let $D := E$, and Proposition 6.1 is proved.

4. If there is some h for which case B holds, then, by (6.9), we have for either $i = 1$ or $i = 2$ that

$$\sum_{\substack{n \in E \\ n | w_i}} \frac{1}{n} \geq \frac{1}{2} \sum_{\substack{n \in E \\ n | w_1 \text{ or } w_2}} \frac{1}{n} > \frac{1}{2}\left(\sum_{n \in E} \frac{1}{n} - \frac{2N^{1-\theta}}{(\log\log N)^2 N}\right) > 1 - O\left(\frac{1}{N^\theta (\log\log N)^2}\right).$$

Without loss of generality, assume that the inequality holds for $i = 1$, and let E^* be those elements of E which divide w_1.

5. Use Proposition 6.2 again, but this time with $J = E^*$, $\alpha = \sum_{n \in E^*} 1/n$, and $\nu = 2/3$, to produce a set D_j satisfying (6.7) and (6.8) with $E = D_j$. From (6.11) we have that $\Sigma(D_j) < \Sigma(E^*) < (1 - e^{-1} + o(1))\log\log N$.

6. Set $C_{j+1} = C_j \setminus D_j$. If $C_{j+1} \leq 8/3$, then STOP; else, increment j by 1 and go back to step 2.

When this procedure terminates, we are either left with a set D from step 3 which proves our Proposition, or we are left with 6 disjoint sets, $D_1, ..., D_6 \subset \mathcal{C}'(N, N^{1+\delta}; \theta)$ satisfying $\sum_{n \in D_i} 1/n \in [2/3 - 1/N, 2/3)$ and

$$(e^{-1} - o(1)) \log\log N < \Sigma(D_i) < (1 - e^{-1} + o(1)) \log\log N. \qquad (6.12)$$

The lower bound for $\Sigma(D_i)$ follows from Lemma 6.3 with $H = D_i$, and the upper bound is as given in step 5.

We claim that there exist three of our sets, D_a, D_b, D_c such that if $L = \mathcal{Q}_{D_a} \cap \mathcal{Q}_{D_b} \cap \mathcal{Q}_{D_c}$, then $\Sigma(L) \gg \log\log N$. For any such triple, we will show that letting $D = D_a \cup D_b \cup D_c$ satisfies the conclusions of Proposition 6.1.

To show that D_a, D_b, D_c exist, let R be the set of prime powers $\leq N^\theta$ which are contained in at least three of the sets $\mathcal{Q}_{D_1}, ..., \mathcal{Q}_{D_6}$. Then, by (6.12),

$$\Sigma(R) > \frac{1}{4}\left(\sum_{i=1}^{6} \Sigma(D_i) \;-\; 2 \sum_{\substack{p^a \leq N^\theta \\ p \text{ prime}}} \frac{1}{p^a}\right) > \frac{1}{4}\left(\frac{6}{e} - 2 - o(1)\right) \log\log N \gg \log\log N$$

Thus, since there $20 = \binom{6}{3}$ triples of sets chosen from $\{D_1, ..., D_6\}$, there is at least one such triple which gives $\Sigma(L) > \Sigma(R)/20 \gg \log\log N$.

Now, letting $D = D_a \cup D_b \cup D_c$ certainly satisfies (6.2). Suppose that the number of elements of D which do not divide any element of I is at most $N^{1-\theta}/(\log\log N)^2$. Then, the hypotheses of Propostion 6.3 hold for $E = D_a, D_b$, and D_c. Case B of Proposition 6.3 cannot hold for $E = D_a$ (or D_b, or D_c), else (6.11) and (6.12) would give us

$$\begin{aligned}\sum_{\substack{q \mid \gcd(w_1, w_2) \\ q \in \mathcal{Q}_{D_a}}} \frac{1}{q} \;&>\; \sum_{\substack{q \mid w_1 \\ q \in \mathcal{Q}_{D_a}}} \frac{1}{q} + \sum_{\substack{q \mid w_2 \\ q \in \mathcal{Q}_{D_a}}} \frac{1}{q} - \sum_{q \in \mathcal{Q}_{D_a}} \frac{1}{q} \\ &>\; \left(\frac{3}{e} - 1 - o(1)\right) \log\log N \gg \log\log N,\end{aligned}$$

which, by Lemma 6.2, would imply that $w_1 = w_2$. Thus, Case A of Proposition 6.3 holds for $E = D_a, D_b$, and D_c: Let W_a, W_b, and W_c be the single integer in I dividing all elements of D_a, D_b, and D_c, respectively, and thus they are all divisible by every element of L. Since $\Sigma(L) \gg \log\log N$, we have, from Lemma 6.2, that $W_a = W_b = W_c = W$, for some $W \in I$. Proposition 6.1 now follows since $\text{lcm}\{n \in D\}|W$.

6.5 Proof of Proposition 6.2

To prove Proposition 6.2 we will need the following lemma.

Lemma 6.4 *Suppose S is a set of integers, all of whose prime power divisors are less than N, which satisfies $\sum_{n\in S} 1/n \geq \rho > \mu$. If $N \gg 1$, then there exists a subset $T \subseteq S$ for which*

$$\sum_{n\in T} \frac{1}{n} > \mu, \quad \text{and} \quad \sum_{\substack{n\in T\\ q|n}} \frac{1}{n} > \frac{\rho-\mu}{2q\log\log N}, \quad \text{for all } q \in \mathcal{Q}_T. \tag{6.13}$$

Proof. We form a chain of subsets $S_0 := S \supset S_1 \supset \cdots \supset T := S_k$ as follows: given S_i, let q_i be the smallest prime power such that

$$\sum_{\substack{n\in S_i\\ q_i|n}} \frac{1}{n} < \frac{\rho-\mu}{2q_i\log\log N},$$

if such q_i exists, and then let $S_{i+1} = S_i \setminus \{n \in S_i : q_i|n\}$. If no such q_i exists, then let $k = i$ and $T = S_i = S_k$. We have that

$$\sum_{n\in T} \frac{1}{n} > \rho - \frac{\rho-\mu}{2\log\log N} \sum_{\substack{p^a\leq N\\ p \text{ prime}}} \frac{1}{p^a} > \mu,$$

for N large enough, since $\sum_{p^a\leq N} 1/p^a < 2\log\log N$.

Proof of Proposition 6.2. We first use Lemma 6.4 with $\rho = \alpha$, $\mu = \nu$, and $S = J$, to produce a set $D_0 = T$ satisfying (6.13). Thus, (6.8) holds for $E = D_0$.

We will construct a chain of subsets $D_0 \supset D_1 \supset D_2 \supset \cdots$, where each set D_j satisfies (6.8) with $E = D_j = D_{j-1} \setminus \{w_j\}$, where w_j is some yet to be chosen element of D_{j-1}. If we can do this then we will eventually reach a set D_k which also satisfies (6.7), since each $w_j \geq N$, and so the Proposition will be proved.

Suppose (6.8) is satisfied for $E = D_{j-1}$, for $j \geq 1$. Take Lemma 6.4 with $S = D_{j-1}$, $\rho = \nu$, and $\mu = \nu/2$, and let w_j be the smallest element of T. Let $q \in \mathcal{Q}_{D_j}$. If $q \nmid w_j$, then

$$\sum_{\substack{n \in D_j \\ q|n}} \frac{1}{n} = \sum_{\substack{n \in D_{j-1} \\ q|n}} \frac{1}{n} > \frac{\min\{\nu, \alpha - \nu\}}{5q \log\log N},$$

by hypothesis. On the other hand, if $q|w_j$, then, by (6.13), we get

$$\sum_{\substack{n \in D_j \\ q|n}} \frac{1}{n} \geq \sum_{\substack{n \in T \\ q|n}} \frac{1}{n} - \frac{1}{w_j} > \frac{\nu}{4q \log\log N} - \frac{1}{N} > \frac{\nu}{5q \log\log N},$$

since $q \leq N^{\theta}$, with $\theta < 1$, and $\nu \gg 1$, and so (6.8) holds for $E = D_j$.

6.6 Proof of Proposition 6.3

Let E_I denote the set of integers in E which divide an integer in I. Then we have, by hypothesis, that $|E_I| > |E| - N^{1-\theta}/(\log\log N)^2$. If $q \in \mathcal{Q}_E$, then

$$\sum_{\substack{n \in E_I \\ q|n}} \frac{1}{n} > \sum_{\substack{n \in E \\ q|n}} \frac{1}{n} - \frac{N^{1-\theta}}{N(\log\log N)^2} \gg \frac{1}{q \log\log N}, \tag{6.14}$$

since $q \leq N^{\theta}$ and E satisfies (6.8). Thus, we have that $\mathcal{Q}_{E_I} = \mathcal{Q}_E$.

We will show at the end of this section that for all $q \in \mathcal{Q}_E$, there exists an integer $qd \in [N^{3/4}, N^{3/4+\theta}]$ such that

$$\sum_{\substack{n \in E_I \\ qd|n}} \frac{1}{n} \gg_{\theta} \frac{1}{qd(\log\log N)^2}, \tag{6.15}$$

where $\omega(d) \leq \omega_0 = \log\log N / \log\log\log\log N$, for N sufficiently large, and all the prime divisors of d are greater than $y := \exp((1/8 - \theta/2)\log N / \log\log N)$.

For now, let us assume that this is true and let qd satisfy (6.15) for a given $q \in \mathcal{Q}_E$. All the elements of E_I which are divisible by qd must divide the same number $n(q) \in I$, since otherwise there are at least two numbers $n_1 < n_2$ in E_I where $qd \mid \gcd(n_1, n_2) \mid (n_2 - n_1)$, which is impossible since $0 < (n_2 - n_1) \leq N^{3/4}$, whereas $qd \geq N^{3/4}$. We will show that as a consequence of this and (6.15),

$$\sum_{p^a | n(q)} \frac{1}{p^a} > \left(\frac{1}{e} - o(1)\right) \log\log N. \tag{6.16}$$

This implies there are at most two distinct values of $n(q)$, for all $q \in \mathcal{Q}_E$: for if there were three prime powers q_1, q_2, q_3 with $n(q_1), n(q_2), n(q_3)$ distinct, then, by Lemma 6.2,

$$\sum_{p^a | \gcd(w_1, w_2)} \frac{1}{p^a} \ll \log\log N,$$

so that, by (6.16),

$$\begin{aligned} \log\log N + O(1) = \sum_{\substack{p^a \leq N \\ p \text{ prime}}} \frac{1}{p^a} > \Sigma(E) &\geq \sum_{i=1}^{3} \sum_{p^a | n(q_i)} \frac{1}{p^a} + O(\log\log\log N) \\ &> (3e^{-1} - o(1)) \log\log N, \end{aligned}$$

which is impossible.

If there is just one value for $n(q)$, for all $q \in \mathcal{Q}_E$, then $w = n(q)$ satisfies Case A of Proposition 6.3: Otherwise, there are two possible values for $n(q)$, call them w_1 and w_2, which satisfy (6.10). The lower bound in (6.11) comes from (6.16). Moreover,

$$\sum_{\substack{q | w_1 \\ q \in \mathcal{Q}_E}} \frac{1}{q} \leq \sum_{p^a \leq N} \frac{1}{p^a} - \sum_{\substack{q | w_2 \\ q \in \mathcal{Q}_E}} \frac{1}{q} + \sum_{p^a | \gcd(w_1, w_2)} \frac{1}{p^a} = (1 - e^{-1} + o(1)) \log\log N,$$

which implies the upper bound in (6.11) (note: the same upper bound holds for w_2), using the Prime Number Theorem, (6.16), and Lemma 6.2, respectively.

If w_1, w_2 fail to satisfy (6.9), then

$$\#\{n \in E_I : n \nmid w_1 \text{ or } w_2\} > \#\{n \in E : n \nmid w_1 \text{ or } w_2\} - \frac{N^{1-\theta}}{(\log\log N)^2} > \frac{N^{1-\theta}}{(\log\log N)^2}.$$

Since there are $\leq N^{3/4}$ integers in I, there must exist an integer $x \in I, x \neq w_1$ or w_2, for which

$$\#\{n \in E_I : n|x\} \gg \frac{N^{1-\theta}}{N^{3/4}(\log\log N)^2} = \frac{N^{1/4-\theta}}{(\log\log N)^2} \tag{6.17}$$

Therefore,

$$\mathrm{lcm}_{n\in E, n|x}\ n \leq \gcd(x, w_1 w_2) \leq \gcd(x, w_1)\gcd(x, w_2) < (x-w_1)(x-w_2) < N^{3/2};$$

but then we have

$$\#\{n \in E\ :\ n|x\} \leq \tau\left(\mathrm{lcm}_{n\in E, n|x}\ n\ \right) \leq \max_{l\leq N^{3/2}} \tau(l) = N^{o(1)},$$

which contradicts (6.17), and so (6.9) follows. Thus, the proof of Proposition 6.3 is complete once we establish (6.15) and (6.16).

To show (6.16), we observe that every integer $m \in F = \{n/qd : n \in E, qd|n\}$ satisfies $\omega(m) \sim \log\log N$, since $\omega(qd) \leq \omega_0 = o(\log\log N)$, and since $E \subseteq \mathcal{C}'(N, N^{1+\delta}; \theta)$. From this and (6.15), F satisfies the hypotheses of Lemma 6.3 with $H = F$. Thus, $\Sigma(F) > (e^{-1} - o(1))\log\log N$, which implies (6.16).

We will now establish (6.15). First, we claim that for every $n \in E$, where $q|n$ and $q \in \mathcal{Q}_E$, there exists a divisor $qd \in [N^{3/4}, N^{3/4+\theta}]$, where $p|d$ implies $p > y$ (though it may not be the case that $\omega(d) \leq \omega_0$). To show this, we construct such a d by adding on prime factors one at a time, until qd is in this interval. There are enough prime factors $> y$ to do this, since for $N \gg_\epsilon 1$ we have

$$\prod_{\substack{p^a \| n/q \\ p>y}} p^a > \frac{n}{q\prod_{\substack{p^a\|n \\ p\leq y}} p^a} > \frac{n}{qy^{\Omega(n)}} > \frac{N}{N^\theta \exp\left(\frac{(1/8-\theta/2)\Omega(n)\log N}{\log\log N}\right)} > N^{3/4},$$

for N sufficiently large, since $\Omega(n) \sim \log\log N$.

If (6.15) fails to hold for all $d \in [N^{3/4}/q, N^{3/4+\theta}/q]$ with $\omega(d) \leq \omega_0$, then we would have by (6.8) and Mertens's Theorem that,

$$\frac{\min\{\nu, \alpha-\nu\}}{5q\log\log N} < \sum_{\substack{n\in E \\ q|n}} \frac{1}{n} < \sum_{\substack{N^{3/4}/q\leq d\leq N^{3/4+\theta}/q \\ p|d \Rightarrow p>y}} \sum_{\substack{n\in E \\ qd|n}} \frac{1}{n}$$

$$< \sum_{\substack{N^{3/4}/q \le d \le N^{3/4+\theta}/q \\ p|d \Rightarrow p>y \\ \omega(d)<\omega_0}} \sum_{\substack{n \in E \\ qd|n}} \frac{1}{n} + \sum_{\substack{d:p|d \Rightarrow y<p<N \\ \omega(d) \ge \omega_0}} \sum_{\substack{m \le N^{1+\delta}/qd \\ (n=mqd)}} \frac{1}{qdm}$$

$$= o\left(\frac{1}{q(\log\log N)^2} \sum_{d:p|d \Rightarrow y<p<N} \frac{1}{d}\right) + O\left(\frac{\log N}{q} \sum_{\substack{d:p|d \Rightarrow y<p<N \\ w(d) \ge \omega_0}} \frac{1}{d}\right). \tag{6.18}$$

Now,

$$\sum_{d:p|d \Rightarrow y<p<N} \frac{1}{d} \le \prod_{\substack{y<p<N \\ p \text{ prime}}} \left(1 - \frac{1}{p}\right)^{-1} \ll \frac{\log N}{\log y} \ll \log\log N,$$

by Mertens's Theorem, and for $k = (\log\log\log N)^3$, we have, again by Mertens's Theorem,

$$\sum_{\substack{d:p|d \Rightarrow y<p<N \\ \omega(d) \ge \omega_0}} \frac{1}{d} \ll \sum_{d:p|d \Rightarrow y<p<N} \frac{k^{\omega(d)-\omega_0}}{d} = \frac{1}{k^{\omega_0}} \prod_{\substack{y<p<N \\ p \text{ prime}}} \left(1 + \frac{k}{p-1}\right)$$

$$= \frac{1}{k^{\omega_0}} \left(\frac{\log N}{\log y}\right)^{k+o(k)} \ll \frac{1}{\log^2 N}.$$

Combining these two applications of Mertens's Theorem with (6.18), we arrive at a contradiction. Thus, there must exist a $d \in [N^{3/4}/q, N^{3/4+\theta}/q]$ satisfying (6.15), with $\omega(d) \le \omega_0 = o(\log\log N)$.

Chapter 7

Conclusion

The methods that we have introduced in Chapters 4, 5, and 6, although successful at solving the problems of Erdős and Graham mentioned in the abstract, have lead to some unsolved questions themselves. For instance, in Chapter 4, we use an exponential sum method to prove that for any integer n and any residue class $l \pmod{n}$, there exist distinct primes $p_1, p_2, ..., p_k < \log^{c+o(1)} n$ such that $1/p_1 + 1/p_2 + \cdots + 1/p_k \equiv l \pmod{n}$, whenver $c \geq 3$. It would be desirable to improve this result so that it works for any $c \geq 1$, and we note that c cannot be taken to be smaller than 1, since the number of subsets of primes $< \log^{1-\epsilon} n$ is at most $2^{\log^{1-\epsilon} n}$, which is smaller than n, the number of choices for l.

In Chapters 5 and 6, we use a different exponential sum method to find unit fraction representations $r = 1/n_1 + \cdots + 1/n_k$, and the method only works if each n_i has all of its prime power divisors $\leq n_i^{1/4-o(1)}$. In Chapter 5 we maneuver around this limitation by combining this method with those from Chapter 4; however, there seems to be no easy way to do this for the results in Chapter 6. Ideally, the methods in Chapters 5 and 6 should be improvable so that each n_i is allowed to have all its prime power divisors $\leq n_i^{1-o(1)}$. Such an improvement would allow one to show the unit fraction positive density conjecture of Erdős and Graham mentioned in the introduction (see [12] and [15]): Given any number $0 < \rho < 1$, and any sequence of density $> \rho$, does there exist a finite subsequence whose sum of reciprocals equals 1? We note that the results in Chapter 6 imply that there exists a ρ_0 so that such a subsequence exists if $0 < \rho_0 < \rho < 1$; however, if $\rho < \rho_0$, then nothing can be said, at present. Because the integers n which are $n^{1-o(1)}$-smooth constitute almost all of the integers, we have that the following, conjectured generalization of the Main Theorem in Chapter 6 would suffice to prove this conjecture of Erdős and Graham:

Conjecture. *Suppose $C \subseteq \mathcal{C}'(N, N^{1+\delta}; \theta)$, where $\theta, \delta > 0$, and $\delta + \theta < 1$. If $N \gg_{\theta,\delta} 1$, and*

$$\sum_{n \in C} \frac{1}{n} > 6,$$

then there exists a subset $S \subset C$ for which $\sum_{n \in S} 1/n = 1$.

One can only hope that some brave soul will attempt this and succeed in the not-too-distant future.

Bibliography

[1] T. Brown, and V. Rödl, *Monochromatic Solutions to Equations with Unit Fractions*, Bull. of the Austral. Mat. Soc., **43** (1991), 387-392.

[2] N. G. de Bruijn, *On the Number of Positive Integers $\leq x$ and Free of Prime Factors $> y$*, Nederl. Akad. Wetensch. Proc. Ser. A **54** (1951), 50-60.

[3] A. B. Chace, H. P. Manning, and R. C. Archibald, *The Rhind Mathematical Papyrus, Volume I*, Math. Assoc. of America, 1927, p. 7, 9.

[4] E. Croot, *On Some Questions of Erdős and Graham about Unit Fractions*, To appear in Mathematika.

[5] ——, *On Unit fractions with Denominators in Short Intervals*, To appear in Acta Arith.

[6] E. Croot, D. Dobbs, J. Friedlander, A. Hetzel, and F. Pappalardi, *Binary Egyptian Fractions*, To appear in J. of Number Theory.

[7] D. R. Curtiss, *On Kellogg's Diophantine Problem*, Amer. Math. Monthly, **29** (1922) 380-387.

[8] Harold Davenport, Multiplicative Number Theory, Second Edition (revised by H. L. Montgomery), Springer-Verlag, New York, 1980.

[9] K. Dickman, *On the Frequency of Numbers Containing Prime Factors of a Certain Relative Magnitude*, Ark. Mat. Astr. Fys. **22** (1930), 1-14.

[10] M. Dunton and R. E. Grimm, *Fibonacci on Egyptian Fractions*, Fibonacci Quart. **4** (1966), 339-354.

[11] C. Elsholtz, *Sums of k Unit Fractions*, 1998, Ph. D. Thesis, Technische Universität Darmstadt.

[12] Paul Erdős and R. L. Graham, *Old and New Problems and Results in Combinatorial Number Theory*, L'Enseignement Mathématique, 1980, pp. 39-40, 103.

[13] R. L. Graham, *On Finite Sums of Unit Fractions*, Proc. London Math. Soc., **14** (1964) 193-207.

[14] ———, *On Finite Sums of Reciprocals of Distinct n-th Powers*, Pac. J. of Math., **14** (1964) 196-197.

[15] Richard K. Guy, *Unsolved Problems in Number Theory, Second Edition*, Springer Verlag, 1994, pp. 158-166.

[16] H. Halberstam and H.-E. Richert, *Sieve Methods* (London Mathematical Society Monographs No. 4), Academic Press, 1974.

[17] G. Hofmeister and P. Stoll, *Note on Egyptian Fractions*, J. Reine Angew. Math., **362** (1985) 141-145.

[18] O. D. Kellogg, *On a Diophantine Problem*, Amer. Math. Monthly, **28** (1921) 300-303.

[19] G. Martin, *Dense Egyptian Fractions*, Trans. Amer. Math. Soc., **351** (1999), 3641-3657.

[20] ———, *Denser Egyptian Fractions*, (Preprint).

[21] Y. Matiyasevich (with foreward by Martin Davis), *Hilbert's Tenth Problem*, MIT Press, 1993, Cambridge, MA.

[22] H. L. Montgomery, *Ten Lectures on the Interface between Analytic Number Theory and Harmonic Analysis*, C.B.M.S. Regional Conference Ser. Math **84**, Amer. Math. Soc, 1994.

[23] L. J. Mordell, *Diophantine Equations*, Academic Press, 1969, London.

[24] Trygve Nagell, *Introduction to Number Theory, Second Edition*, Chelsea Publishing Company, 1964, pp. 54-67

[25] Leonardo Pisano, *Scritti, Vol. 1*, B. Boncompagni, 1857, Rome.

[26] G. Robins and C. Shute, *The Rhind Mathematical Papyrus*, Dover Publications, New York, 1987, pp. 8-18.

[27] J. B. Rosser and Lowell Shoenfeld, *Approximate Formulas for Some Functions of Prime Numbers*, Illinois J. Math. **70** (1962), p. 64-94.

[28] Shen Zun, *On the Diophantine Equation* $\sum_{i=0}^{k} 1/x_i = a/n$, Chinese Annals of Math., **7B** (1986) 213-220.

[29] W. Sierpiński, *Sur les Décompositions de Nombres Rationnels en Fractions Primaries*, Mathesis, **65** (1956) 16-32.

[30] R. C. Vaughan, *On a Problem of Erdős, Straus, and Schinzel*, Mathematika, **17** (1970) 193-198.

[31] C. Viola, *On the Diophantine Equation* $\prod_0^k x_i - \sum_0^k x_i = n$ *and* $\sum_0^k 1/x_i = a/n$, Acta Arith., **22** (1973) 339-352.

[32] M. Vose, *Egyptian Fractions*, Bull. London Math. Soc, **17** (1985) 21-24.

[33] H. Yokota, *On Number of Integers Representable as a Sum of Unit Fractions*, Canad. Math. Bull., **33** (1990) 235-241.

[34] ———, *On a Problem of Erdős and Graham*, J. Number Theory, **39** (1991) 327-338.

[35] ———, *On Number of Integers Representable as a Sum of Unit Fractions, II*, J. Number Theory, **67** (1997) 162-169.

[36] ———, *The Largest Integer Expressible as a Sum of Reciprocals of Integers*, J. Number Theory, **76** (1999) 206-216.