Stefan PLANITZER

Sums of unit fractions, Romanov type problems and Sequences with Property P

DOCTORAL THESIS

to achieve the university degree of

Doktor der technischen Wissenschaften

submitted to

Graz University of Technology

Supervisor:

Assoc.Prof. Dipl.-Math. Dr.rer.nat.habil. Christian ELSHOLTZ

Institute of Analysis and Number Theory

Graz, June 2018

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present doctoral thesis.

Date

Signature

Abstract

The present thesis covers results concerning three different topics in number theory. In the first part we present new results on the number of representations of positive rational numbers as sums of unit fractions. In particular we improve the best known upper bounds on the number of positive integer solutions (a_1, a_2, a_3) of the Erdős-Straus equation $\frac{4}{n} = \frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3}$, for given $n \in \mathbb{N}$. Furthermore, we improve upper bounds on the number of representations of general positive rational numbers as a sum of k unit fractions. For given $m \in \mathbb{N}$ we prove lower bounds on the number of representations of $\frac{m}{n}$ as a sum of three unit fractions for n in different subsets of the positive integers.

The second part covers two problems of Romanov type. Here we prove that the lower density of integers of the forms $p + 2^{2^k} + m!$ and $p + 2^{2^k} + 2^q$ is positive, where k, m are non-negative integers and p, q are primes. Furthermore, we show that also the lower density of odd integers not of these forms is positive.

Finally we deal with sequences with Property P. These sequences of positive integers are characterized by the property, that no element of the sequence divides the sum of two larger ones. We improve a construction by Erdős and Sárközy and give an example of a sequence S with Property P whose counting function $S(x)$ is in a sense large for all $x > 0$.

Kurzfassung

Die vorliegende Arbeit enthält Resultate aus drei verschiedenen Teilbereichen der Zahlentheorie. Im ersten Teil präsentieren wir neue Resultate zur Anzahl der Darstellungen von positiven rationalen Zahlen als Summe von Stammbrüchen. Insbesondere verbessern wir die derzeit besten bekannten oberen Schranken für die Anzahl der Lösungen der Erdős-Straus Gleichung $\frac{4}{n} = \frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3}$ in natürlichen Zahlen (a_1, a_2, a_3), wobei $n \in \mathbb{N}$ gegeben ist. Darüber hinaus verbessern wir obere Schranken für die Anzahl der Darstellungen allgemeiner positiver rationaler Zahlen als Summe von k Stammbrüchen. Für eine gegebene Zahl $m \in \mathbb{N}$ beweisen wir untere Schranken für die Anzahl der Darstellungen von rationalen Zahlen der Form $\frac{m}{n}$ als Summe von drei Stammbrüchen, wobei n jeweils in verschiedenen Teilmengen der natürlichen Zahlen liegt.

Der zweite Teil umfasst Probleme vom Romanov Typ. Wir beweisen, dass die untere Dichte von natürlichen Zahlen der Formen $p + 2^{2^k} + m!$ und $p + 2^{2^k} + 2^q$ positiv ist, wobei k und m natürliche Zahlen und p und q Primzahlen sind. Außerdem beweisen wir, dass auch die untere Dichte jener ungeraden natürlichen Zahlen, die nicht von der entsprechenden Form sind, positiv ist.

Im dritten Teil beschäftigen wir uns mit Folgen mit Property P. Diese Folgen natürlicher Zahlen werden durch die Eigenschaft charakterisiert, dass kein Element der Folge die Summe zweier größerer Elemente teilt. Wir verbessern eine Konstruktion von Erdős und Sárközy und geben ein Beispiel einer Folge S mit Property P an, deren Zählfunktion $S(x)$ in gewisser Weise groß für alle $x > 0$ ist.

List of publications

C. Elsholtz, F. Luca and S. Planitzer, *Romanov type problems*, Ramanujan J. (2018). Available online at https://doi.org/10.1007/s11139-017-9972-8.

C. Elsholtz and S. Planitzer, *The number of solutions of the Erdős-Straus Equation and sums of k unit fractions*, submitted. Available online at https://arxiv.org/abs/1805.02945.

C. Elsholtz and S. Planitzer, *On Erdős and Sárközy's sequences with Property P*, Monatsh. Math. **182** (2017), no. 3, 565–575. Available online at https://doi.org/10.1007/s00605-016-0995-9.

M. G. Madritsch and S. Planitzer, *Romanov's Theorem in Number Fields*, submitted. Available online at https://arxiv.org/abs/1512.04869.

Acknowledgements

First and most of all I would like to thank my advisor Christian Elsholtz, for his time, guidance and encouragement. Working with him for the last years was a pleasure and a privilege.

I am grateful, that as a Ph.D. student I could be part of the Doctoral Program Discrete Mathematics, which was funded by the FWF (project W1230). In particular I thank the head of project, Prof. Wolfgang Woess, the DK coordinators Johannes Cuno, Dijana Kreso and Mario Weitzer as well as the secretaries of the Institute of Analysis and Number Theory, Hermine Panzenböck and Irene Pfeifer-Wilfinger, for taking care of all administrative matters. Furthermore, I would like to thank my mentors within this program, Daniel Krenn and Prof. Robert Tichy, for useful advice on many occasions.

The Institut Élie Cartan de Lorraine at Université de Lorraine in Nancy and the Institute of Mathematics at University of Rostock hosted me for two extended research stays. I want to thank these institutions and especially Manfred Madritsch and Jan-Christoph Schlage-Puchta for their hospitality.

In addition to some of the people just mentioned, I had the chance to work with Florian Luca during one of his visits at Graz University of Technology. I would like to thank him and my other co-authors for sharing their knowledge and giving me the opportunity to work with them.

It is my fortune that I have friends who from time to time insist on doing things apart from maths. Among many others I thank Anna, Tina, Gerhard, Markus and Niels for keeping our cinematic group alive and active, Petra and Vera for being the best neighbors you could imagine and Birgit and Verena for amazing experiences all around the world.

My family, in particular my parents Patrizia and Günter, my grandparents Else, Kurt, Olga and Wolfgang, my sister Iris and my godchildren Milena and Samuel, I thank for their love and support over all these years.

Finally, my thanks to Thomas, for everything.

Contents

1. Introduction

This thesis contains results on three different problems in number theory: sums of unit fractions in Chapter 2, Romanov type problems in Chapter 3 and sequences with Property P in Chapter 4. All these chapters consist of a scientific paper on the corresponding subject.

The overview of previous results concerning these problems, which we give in this introduction, does not aim at completeness. In fact we rather focus on those results which have a connection to the topics discussed in the following chapters. Furthermore, we briefly point out the main ideas that we will apply later.

1.1. Unit fractions

We call positive rational numbers with a representation of the form $\frac{1}{n}$, $n \in \mathbb{N}$, i.e. a representation where the numerator is 1, a unit fraction. Our main focus is on representations of positive rational numbers as sums of k unit fractions. This leads to Diophantine equations of the form

$$\frac{m}{n} = \sum_{i=1}^{k} \frac{1}{a_i}, \tag{1.1}$$

where m, n and a_i, $1 \leq i \leq k$, are positive integers.

There are a lot of questions connected to equation (1.1). Given $m, n \in \mathbb{N}$, we could for example ask whether for some $k \in \mathbb{N}$ a solution in $(a_1, \ldots, a_k)$ exists. While this question may be trivially answered in a positive way via the representation $\frac{m}{n} = \sum_{i=1}^{m} \frac{1}{n}$, the answer if we additionally require the a_i to be pairwise distinct is less obvious. Indeed by results of Fibonacci (for an English translation of the corresponding parts of his work see for example [17]) and Sylvester [72] we know that the answer also with this additional requirement is positive for positive rational numbers less than 1[1]. Their work connects solutions of equation (1.1) with algorithmic aspects. They observe that a greedy

[1] The method suggested by Fibonacci and Sylvester to find such solutions may be extended to work for general positive rational numbers (see [71, p. 201: Theorem 2]).

approach in the sense of iteratively subtracting the largest unit fraction $\frac{1}{a_i}$, such that the remainder $\frac{m}{n} - \frac{1}{a_1} - \cdots - \frac{1}{a_i}$ is non-negative, produces a representation of the required form after finitely many steps (i.e. after a finite number of subtractions the remainder will be 0).

An important thread of recent research concerns the number of solutions of equation (1.1). Here we adopt the notation from [9] and define $f_k(m,n)$ to denote the number of these solutions with $a_1 \leq a_2 \leq \ldots \leq a_k$, where we consider m, n and k to be fixed. For $k = 3$ and $m = 4$ this leads to a famous conjecture by Erdős and Straus (see e.g. the English summary of [28]), stating that for any positive integer $n \geq 2$ there exists at least one solution of the equation

$$\frac{4}{n} = \frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3} \tag{1.2}$$

in positive integers a_1, a_2 and a_3. This conjecture is still open. A well known partial result (see [62, p. 287f.]) is that any exceptional integer n is 1 mod 24, and more precisely in one of the residue classes

$$1, 121, 169, 289, 361, 529 \bmod 840.$$

Furthermore, upper bounds on the number of solutions of equation (1.2) are known. The following is a Corollary to [9, Theorem 2].

Theorem 1.1 (Browning, Elsholtz (2011))**.** *For any $\epsilon > 0$, we have*

$$f_3(4,n) \ll_\epsilon n^{2/3+\epsilon}.$$

For prime denominators this was improved to the following bound in [25, Proposition 1.7].

Theorem 1.2 (Elsholtz, Tao (2013))**.** *For any prime p and any $\epsilon > 0$, we have*

$$f_3(4,p) \ll_\epsilon p^{3/5+\epsilon}.$$

Browning and Elsholtz [9, Theorem 3] also proved the following upper bounds on representations as sums of k unit fractions for general positive rational numbers $\frac{m}{n}$ and $k > 3$.

Theorem 1.3 (Browning, Elsholtz (2011)). *For any $\epsilon > 0$, we have*

$$f_4(m,n) \ll_\epsilon n^\epsilon \left(\left(\frac{n}{m}\right)^{5/3} + \frac{n^{4/3}}{m^{2/3}} \right),$$

and for $k \geq 5$

$$f_k(m,n) \ll_\epsilon (kn)^\epsilon \left(\frac{k^{4/3}n^2}{m}\right)^{5/3 \cdot 2^{k-5}}.$$

Finally we conclude this small survey of known results by mentioning one more special case of equation (1.1) that received some attention also recently. This concerns the number of representations of 1 as a sum of k unit fractions. The best known upper bound is again due to Browning and Elsholtz [9, Theorem 4].

Theorem 1.4 (Browning, Elsholtz (2011)). *Let $\epsilon > 0$, then there exists $k(\epsilon)$ such that for $k \geq k(\epsilon)$*

$$f_k(1,1) < c_0^{(5/12+\epsilon)2^{k-1}},$$

where $c_0 = 1.264\ldots$[2].

Quite recently Konyagin [51] proved the following double exponential lower bound, even for the subclass of representations of 1 as a sum of k distinct unit fractions.

Theorem 1.5 (Konyagin (2014)). *As $k \to \infty$, the number of representations of* 1 *as a sum of k distinct unit fractions is bounded from below by*

$$\exp\left(\exp\left(\left(\frac{(\log 2)(\log 3)}{3} + o(1)\right)\frac{k}{\log k}\right)\right).$$

In Chapter 2 we improve some of these results. First we apply methods developed by Elsholtz and Tao [25] to prove that the bound in Theorem 1.2 does not only hold for prime denominators, but for arbitrary ones.

In the case of prime denominators $n = p$ two types of solutions of equation (1.2) can occur: those where exactly two of the denominators of the unit fractions are divisible by p and those where this is the case for exactly one of the denominators, i.e.

$$\frac{4}{p} = \frac{1}{pt_1} + \frac{1}{pt_2} + \frac{1}{t_3}, \ \ \gcd(p, t_3) = 1,$$

or

$$\frac{4}{p} = \frac{1}{pt_1} + \frac{1}{t_2} + \frac{1}{t_3}, \ \ \gcd(p, t_2t_3) = 1.$$

[2] For a proper definition of the constant c_0 see Definition 2.1 below.

Elsholtz and Tao found suitable parametrizations for both types of solutions. They get their upper bound by showing that the number of choices for the corresponding parameters is not too large.

For general denominators n in (1.2) more than two types of solutions can occur. In Chapter 2 we will work with *patterns* (n_1, n_2, n_3) of solutions, which refer to solutions of the type

$$\frac{4}{n} = \frac{1}{n_1 t_1} + \frac{1}{n_2 t_2} + \frac{1}{n_3 t_3},$$

where $n_i | n$ and $\gcd\left(t_i, \frac{n}{n_i}\right) = 1$ for $i \in \{1, 2, 3\}$. We use the concept of relative greatest common divisors to find a parametrization that works for any of these patterns. Since the number of patterns is not too large, also in this case it will suffice to prove upper bounds for the number of choices for the parameters.

As was already the case for the results of Browning and Elsholtz and those of Elsholtz and Tao, the proof of this upper bound is very constructive. This is why the bound for the number of solutions we establish, closely corresponds to an upper bound for the running time of an algorithm which enumerates all these solutions.

Next we improve the upper bound in Theorem 1.3 for representations as sums of k unit fractions for $k > 3$. To do so we apply a lifting method developed by Browning and Elsholtz [9]. The improvement stems from additionally considering a parametrization for the solutions of equations of the form

$$\frac{4}{n} = \frac{1}{n_1 t_1} + \frac{1}{n_2 t_2} + \frac{1}{n_3 t_3} + \frac{1}{n_4 t_4}.$$

Also in this case we get this parametrization from the relative greatest common divisors of the integers t_1, t_2, t_3 and t_4. The better upper bound we derive for sums of k unit fractions in the general case also leads to an improvement of the bound in Theorem 1.4 for the special case of representations of 1 in this form (see Corollary 2.6 in Chapter 2 below).

The Erdős-Straus conjecture shows that already for the number of representations of a positive rational number $\frac{m}{n}$ as a sum of three unit fractions, good lower bounds for all denominators n can be hard to achieve. In Chapter 2 we prove certain lower bounds for the number of representations of rational numbers of the form $\frac{m}{n}$ as a sum of three unit fractions, where the denominator n is either in an infinite subset of the positive integers or the primes, or in a subset of density one within the positive integers. One of the results we will prove in this direction (see Remark 2.18 in Chapter 2) is the existence of infinitely many primes $p \equiv 1 \bmod 4$ such that $f_3(4, p) \gg \exp\left((0.1444 + o(1))\frac{\log p}{\log \log p}\right)$.

1.2. Romanov's theorem

A famous problem in number theory concerns the set of odd integers which are sums of primes and powers of 2. A conjecture of de Polignac [65] stated that any odd integer is the sum of a prime and a power of 2. In a correction concerning this conjecture he mentions a letter of Euler [37] to Goldbach, in which Euler states that 959 is not the sum of a prime and a power of 2. The smallest counter example to de Polignac's conjecture is 127 which in contrast to 959 is a prime.

In view of the existence of prime as well as non-prime counter examples it seems reasonable to take a big step back and ask whether it is even true, that the set of positive integers with a representation as the sum of a prime and a power of 2 has positive lower density. This question was answered in the positive by Romanov [66, Satz II]:

Theorem 1.6 (Romanov (1934))**.** *Given an integer $a \geq 2$, there exists a constant $\beta_a > 0$, depending only on a, such that the lower density of integers which are sums of a prime and a power of a is at least β_a.*

Here we define the *lower density* of a subset A of the positive integers by

$$\liminf_{x\to\infty} \frac{A(x)}{x},$$

where $A(x) := \sum_{\substack{a\leq x\\ a\in A}} 1$, as usual, denotes the counting function of A. By exchanging the limit inferior with the limit superior, we get what we call the *upper density* of A. If lower- and upper density coincide we simply speak of the *density* of the set A.

More recently some effort was put into determining lower bounds on the constant β_2 in Theorem 1.6 (see for example [13, 43, 44, 64]). The current record is held by Elsholtz and Schlage-Puchta [24], who proved that $\beta_2 \geq 0.107648$.

In Chapter 3 we deal with a variant of this problem. We will consider the representation of integers as the sum of a prime, an iterated power of 2 and either a factorial or a prime power of 2. A general question behind Theorem 1.6 and the research in Chapter 3 is the following one. Given a set $A \subseteq \mathbb{N}$ with $A(x) \sim c_A \log x$ for some positive constant c_A depending on A . Is it true, that the sum-set

$$\mathbb{P} + A := \{p + a : p \in \mathbb{P}, a \in A\}$$

has positive lower density in $\mathbb{N}$? In view of the prime number theorem there are enough combinations of primes and elements from the set A such that the answer to this question could be ‘yes’.

When proving results of this type, following the method employed by Romanov [66] in his proof of Theorem 1.6 could lead to success. We summarize the basic ideas behind his method in the following.

For a given set $A \subset \mathbb{N}$ with $A(x) \sim c_A \log x$ we want to prove that the sum-set $\mathbb{P} + A$ has positive lower density. For any $\epsilon > 0$ and sufficiently large x there are at least $(c_A - \epsilon)x$ sums of a prime and an element of A less than x. Hence, informally speaking, this can only go wrong if we have lots of representations as a sum of a prime and an element of A for a small set of integers and only few for the rest.

The study of the lower density of the set $\mathbb{P} + A$ is therefore linked to the study of the corresponding representations function

$$r(n) := \left|\{(p, a) : p \in \mathbb{P}, a \in A, n = p + a\}\right|,$$

and the associated indicator function

$$\mathbb{1}_{\mathbb{P}+A}(n) = \begin{cases} 1, & \text{if } r(n) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

An application of the Cauchy-Schwarz inequality immediately yields

$$\sum_{n \leq x} \mathbb{1}_{\mathbb{P}+A}(n) \geq \frac{\left(\sum_{n \leq x} r(n)\right)^2}{\sum_{n \leq x} r(n)^2}. \tag{1.3}$$

Next we derive a lower bound of the form cx for the sum on the left hand side in (1.3), where c is some positive constant. This reduces to finding a lower bound of order x for $\sum_{n \leq x} r(n)$ and an upper bound of the same order for $\sum_{n \leq x} r(n)^2$.

The lower bound for $\sum_{n \leq x} r(n)$ is typically easier to find. For sums of primes and powers of 2 it suffices for example to bound the representations function $r(n)$ from below by the number of representations where any of the two summands is bounded from above by $\frac{x}{2}$.

Bounding the sum $\sum_{n \leq x} r(n)^2$ is usually the harder task. The squared values $r(n)^2$ of the representation function can be interpreted as the number of pairs of representations of n as the sum of a prime and an element from A, i.e.

$$r(n)^2 = \left|\{(p_1, p_2, a_1, a_2) : p_1, p_2 \in \mathbb{P}, a_1, a_2 \in A, p_1 + a_1 = p_2 + a_2 = n\}\right|. \tag{1.4}$$

From this we see that if we consider a_1 and a_2 to be fixed summing over $r(n)^2$ means counting pairs of primes with a fixed difference. Classical results from sieve theory are used to bound the number of these prime pairs.

In Chapter 3 we will apply Romanov's method to two variants of the original problem. We will consider sums of primes, iterated powers of 2, i.e. integers of the form 2^{2^k}, and either a factorial $m!$ or a power of 2 with a prime exponent.

In view of the prime number theorem and by counting iterated powers of two and factorials less than x, in both cases we get a lower bound of order

$$\frac{x}{\log x} \cdot \log\log x \cdot \frac{\log x}{\log\log x} \gg x$$

for the number of choices for the three summands, if we restrict any of them to be at most $\frac{x}{3}$. Hence it could be that the sets of integers which are of the corresponding two forms have positive lower density. This can only be true, if for the sets

$$A_1 = \left\{n \in \mathbb{N} : n = 2^{2^k} + m!, k, m \in \mathbb{N}_0\right\},$$
$$A_2 = \left\{n \in \mathbb{N} : n = 2^{2^k} + 2^q, k \in \mathbb{N}, q \in \mathbb{P}\right\},$$

we have that $A_1(x) \gg \log x$ and $A_2(x) \gg \log x$. While for A_2 this lower bound follows essentially from the uniqueness of the binary representation of a positive integer, proving $A_1 \gg \log x$ requires more work. In particular we need to make sure that a positive integer n does not have too many representations in the form $n = 2^{2^k} + m!$, $k, m \in \mathbb{N}_0$. This follows from Theorems 3.7 and 3.8 in Chapter 3 where we determine all solutions of the equation $2^{x_1} + y_1! = 2^{x_2} + y_2!$ in non-negative integers x_1, x_2, y_1 and y_2.

The idea for considering these variants of Romanov's problem also was that Romanov's method depends to some degree on the periodic behavior of 2^k modulo odd integers. By replacing 2^k with iterated powers of two we destroy some of this regularity. In particular our results show that the general ideas behind Romanov's method also work for sums of primes and sets of integers exhibiting less periodic behavior than the set of powers of a fixed integer does.

Coming back to sums of primes and powers of 2, just knowing Theorem 1.6 it could still be that essentially almost all odd integers are of this form. This would imply that the density of sums of primes and powers of 2 is $\frac{1}{2}$. A well known result which was established independently by Erdős [29] and van der Corput [14] proves this wrong.

Theorem 1.7 (Erdős (1950), van der Corput (1950)). *The lower density of odd integers not of the form $p + 2^k$, $p \in \mathbb{P}$, $k \in \mathbb{N}$, is positive.*

For this purpose Erdős [29] invented the method of covering congruences to construct a full arithmetic progression of integers which have no representation as the sum of a prime and a power of 2. A covering congruence is a set of residue classes $(a_i \bmod m_i)_{i=1}^{l}$ such that

$$\mathbb{N}_0 \subset \bigcup_{i=1}^{l} \{a_i + jm_i : j \geq 0\}.$$

The following observation is at the heart of Erdős' argument: Suppose that we have a system of covering congruences such that for $1 \leq i \leq l$ we can find distinct primes $p_1, \ldots, p_l$ where p_i divides $2^{m_i} - 1$ (by Zsigmondy's Theorem [76] this is in particular possible if all moduli are at least 2, pairwise distinct and different from 6). Take the intersection of the arithmetic progressions $2^{a_i} \bmod p_i$, $1 \leq i \leq l$, and suppose that the integer n is in this intersection. Then by construction, the difference of n and any power of 2 will always be divisible by one of the primes p_i, $1 \leq i \leq l$.

Erdős gave an example of a system of congruences satisfying the above restrictions which with the previous argument is enough to prove Theorem 1.7. Furthermore, Erdős' system can be extended to rule out integers of the form $p_i + 2^k$, $1 \leq i \leq l$, so that we retrieve a full arithmetic progression of integers not of the form $p + 2^k$.

In Chapter 3 we use similar arguments to prove that the lower density of odd integers which have no representation of the forms $p + 2^{2^k} + m!$ and $p + 2^{2^k} + 2^q$, for $k, m \in \mathbb{N}_0$ and $p, q \in \mathbb{P}$, is positive. Indeed, the lower densities are larger than $\frac{1}{4}$ and $\frac{1}{6}$, i.e. much larger than is known in the original problem for integers not of the form $p + 2^k$ (Habsieger and Roblot [43] prove a lower bound of 0.00905 in this case).

1.3. Sequences with Property P

The last part, i.e. Chapter 4, is about a special kind of sequences of positive integers. In particular we worked on sequences with Property P, which were introduced by Erdős and Sárközy in [34]. The following definition captures the concept of those sequences.

Definition 1.8 (Sequences with Property P)**.** *Let $(a_i)_{i\in\mathbb{N}}$ be an increasing sequence of positive integers, then $(a_i)_{i\in\mathbb{N}}$ has Property P, if a_i does not divide $a_j + a_k$, for all $i < j \leq k$.*

Erdős and Sárközy [34] were primarily interested in questions concerning the density of infinite sequences with Property P. They proved that the density of any infinite sequence with Property P exists and that it is the same for all these sequences.

Theorem 1.9 (Erdős and Sárközy (1970))**.** *Any infinite sequence with Property P has density* 0*.*

Theorem 1.9 is interesting, since it was a priori not clear, whether the density of a sequence with Property P always exists. In particular this should be compared with results on primitive sequences which were a starting point for the investigation of sequences with Property P (an overview of classical results concerning primitive sequences may be found in the book of Halberstam and Roth [45, Chapter V]).

Definition 1.10 (Primitive sequence)**.** *The sequence* $(a_i)_{i\in\mathbb{N}}$ *is called a primitive sequence, if* a_i *does not divide* a_j *for* $i \neq j$*.*

Questions concerning the density of primitive sequences are well studied. As pointed out in [45, p. 244: Theorem 1] the upper density of a primitive sequence is bounded from above by $\frac{1}{2}$. This is simply due to the fact that the greatest odd divisors of the elements in a primitive sequence need to be pairwise distinct. Besicovitch [7] proved that the upper density of a primitive sequence can be arbitrarily close to this upper bound in the following sense.

Theorem 1.11 (Besicovitch (1935))**.** *For a given* $0 < \epsilon < \frac{1}{4}$ *there exists a primitive sequence whose upper density is larger than* $\frac{1}{2} - \epsilon$*.*

The following result by Erdős [26] on the other hand determines the lower density of any primitive sequence.

Theorem 1.12 (Erdős (1935))**.** *Every primitive sequence has lower density* 0*.*

Theorems 1.11 and 1.12 show that, in contrast to sequences with Property P, for primitive sequences the density does not necessarily exist. In particular, for a primitive sequence A the sequence $\left(\frac{A(n)}{n}\right)_{n\in\mathbb{N}}$ can be highly oscillating. Nonetheless, if the density of a primitive sequence exists, it has to be 0.

We note that in the case when any two elements of a sequence with Property P are coprime we know more. Improving a result of Schoen [70], Baier [3] proved the following.

Theorem 1.13 (Baier (2004))**.** *Let* A *be a sequence with Property P consisting of pairwise coprime integers. Then for any given* $\epsilon > 0$*, there are infinitely many* $x \in \mathbb{N}$ *such that*

$$A(x) < (3+\epsilon)\frac{x^{2/3}}{\log x}.$$

In Chapter 4 we deal with the question of how large the counting function of a sequence with Property P can be, which may be studied from different points of view, two of them being:

1. Find a sequence with Property P whose counting function $A(x_\nu)$ is large *for an infinite sequence* $(x_\nu)_{\nu\in\mathbb{N}}$ tending to infinity.

2. Find a sequence with Property P whose counting function $A(x)$ is large *for all* x.

The question looked at from the first of these points of view was answered in the following way by Erdős and Sárközy [34]. They proved that for any increasing function $f(x)$, there exists a sequence A with Property P such that $A(x_\nu) > \frac{x_\nu}{f(x_\nu)}$, for a sequence $(x_\nu)_{\nu\in\mathbb{N}}$ tending to infinity. In view of Theorem 1.9 this is optimal.

Apart from this, Erdős and Sárközy [34] also provide an example of a sequence A with Property P whose counting function $A(x)$ is large *for all* x. They observe, that it is possible to choose A to be the sequence of squares of the primes in the residue class $3 \bmod 4$. The reason why this works is rather simple. It is a well known fact (see for example [47, Theorem 82]) that -1 is a quadratic non-residue modulo any prime in the residue class $3 \bmod 4$. This means that there is no $0 \leq x < p$ such that $x^2 \equiv -1 \bmod p$ for those primes.

Now suppose that there exist primes $p_1 < p_2 \leq p_3$, all of them in the residue class $3 \bmod 4$, such that p_1^2 divides $p_2^2 + p_3^2$. This would in particular imply that

$$p_2^2 \equiv -p_3^2 \bmod p_1.$$

Since $\gcd(p_1, p_3) = 1$, p_3^{-1} exists modulo p_1 and we have $(p_2 p_3^{-1})^2 \equiv -1 \bmod p_1$, a contradiction to the fact that -1 is a quadratic non-residue modulo p_1. This shows that the set of squares of primes $p \equiv 3 \bmod 4$ indeed gives rise to a sequence A with Property P. By the prime number theorem for arithmetic progressions the counting function of this sequence asymptotically behaves like $A(x) \sim \frac{\sqrt{x}}{\log x}$.

In Chapter 4 we improve on this construction by applying the following ideas:

1. Use multiple sets with Property P: We construct infinitely many sets $S_i \subseteq \mathbb{N}$, each with Property P. The basic idea is to choose S_i to be the set of squares of integers with exactly i distinct prime factors $p \equiv 3 \bmod 4$ and no other ones. Given three integers $n_1, n_2, n_3 \in S_i$ with $n_1 < n_2 \leq n_3$, the fact that any of these integers has the exact same number of prime divisors ensures the existence of a prime $p_1 \equiv 3 \bmod 4$ which divides n_1 but does not divide n_2. If the sum $n_2 + n_3$ would be divisible by n_1 it would in particular be divisible by p_1. Hence the reason why the sets S_i have Property P is similar to the reason why the Erdős-Sárközy sequence works.

2. Use indicator factors: We want to consider the set

$$S := \bigcup_{i=1}^{\infty} S_i.$$

The problem is, that even though all the sets S_i constructed in the previous step have Property P, their union does not necessarily have this property, as a_i could divide the sum $a_j + a_k$ if a_i, a_j and a_k are in different sets S_i, S_j and S_k. To fix this we equip every set S_i with a unique indicator factor. More specifically for any set S_i there will be exactly one prime q_i which appears with an even exponent larger than 2 in the prime factorization of all $s \in S_i$. This will imply that a_i can not divide $a_j + a_k$, with a_i, a_j and a_k from different sets S_i, S_j and S_k.

3. The counting function $S(x)$: Finally, we need to determine a lower bound on the counting function $S(x)$. For different x different sets S_i will yield the main contribution to $S(x)$. To see which sets S_i we need to consider, we need to know how many distinct prime factors $p \equiv 3 \bmod 4$ we can expect for an arbitrary positive integer less than x. Results like those in [73, p. 434: eq. (3.38)] show that we should expect most integers $n \leq x$ having their number of prime factors of the form $p \equiv 3 \bmod 4$ in an interval of size $\mathcal{O}(\sqrt{\log\log x})$ centered at $\frac{\log\log x}{2}$.

2. Unit Fractions

This chapter contains an article, which is joint work with Christian Elsholtz. Apart from minor changes, mostly in typesetting, the article below is identical with the version on the arXiv [23].

The number of solutions of the Erdős-Straus Equation and sums of k unit fractions

Christian Elsholtz and Stefan Planitzer

Abstract. We prove new upper bounds for the number of representations of an arbitrary rational number as a sum of three unit fractions. In particular, for fixed m there are at most $\mathcal{O}_\epsilon(n^{3/5+\epsilon})$ solutions of $\frac{m}{n} = \frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3}$. This improves upon a result of Browning and Elsholtz (2011) and extends a result of Elsholtz and Tao (2013) who proved this when $m = 4$ and n is a prime. Moreover, there exists an algorithm finding all solutions in expected running time $\mathcal{O}_\epsilon\left(n^\epsilon \left(\frac{n^3}{m^2}\right)^{1/5}\right)$, for any $\epsilon > 0$. We also improve a bound on the maximum number of representations of a rational number as a sum of k unit fractions. Furthermore, we also improve lower bounds. In particular we prove that for given $m \in \mathbb{N}$ in every reduced residue class $e \bmod f$ there exist infinitely many primes p such that the number of solutions of the equation $\frac{m}{p} = \frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3}$ is $\gg_{f,m} \exp\left(\left(\frac{5\log 2}{12\operatorname{lcm}(m,f)} + o_{f,m}(1)\right)\frac{\log p}{\log\log p}\right)$. Previously the best known lower bound of this type was of order $(\log p)^{0.549}$.

We consider the problem of finding upper bounds for the number of solutions in positive integers a_1, a_2 and a_3 of equations of the form

$$\frac{m}{n} = \frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3} \tag{2.1}$$

where $m, n \in \mathbb{N}$ are fixed. In the case when $m = 4$ we call equation (2.1) Erdős-Straus equation. The Erdős-Straus conjecture states that this equation has at least one solution for any $n > 1$ (see [25] and [42, D11] for classical results concerning the Erdős-Straus equation and several related problems, as well as [41] for a survey of the work of Erdős on egyptian fractions). Also the more general equation

$$\frac{m}{n} = \sum_{i=1}^{k} \frac{1}{a_i}, \tag{2.2}$$

for $m, n \in \mathbb{N}$ fixed and $a_1, \ldots, a_k \in \mathbb{N}$ received some attention. Browning and Elsholtz [9] found upper bounds for the number of solutions of (2.2). For the special case $m = n = 1$ they were able to improve a result of Sándor [69] and proved that there are at most $c_0^{(5/24+\epsilon)2^k}$ representations of 1 as a sum of k unit fractions, for any $\epsilon > 0$ and sufficiently large k. Here c_0 is as in the following Definition (for a proof of the value of c_0 given in Definition 2.1 see [38]).

Definition 2.1. *We define the constant c_0 as*

$$c_0 = \lim_{n\to\infty} u_n^{2^{-n}} = 1.264\ldots,$$

where $(u_n)_{n\in\mathbb{N}}$ is the sequence of positive integers defined by $u_1 = 1$ and $u_{n+1} = u_n(u_n + 1)$.

On the other hand Konyagin [51] proved a lower bound of order

$$\exp\left(\exp\left(\left(\frac{(\log 2)(\log 3)}{3} + o(1)\right)\frac{k}{\log k}\right)\right)$$

for the number of these representations with distinct denominators. While the Erdős-Straus conjecture is about representing certain rational numbers as a sum of just three unit fractions, Martin [57] worked on representations of positive rationals as sums of many unit fractions. In particular he proved that every positive rational number r has a representation of the form $r = \sum_{s\in S} \frac{1}{s}$, where the set S contains a positive proportion of the integers less than any sufficiently large real number x.

Chen et.al. [12] dealt with representations of 1 as a sum of k distinct unit fractions where the denominators satisfy certain restrictions (like all of them being odd). Several results on representations of rational numbers as a sum of unit fractions with restrictions on the denominators can be found in the work of Graham [39–41]. Elsholtz [20] proved a lower bound of similar order as the one of Konyagin for the number of representations of 1 as a sum of k distinct unit fractions with odd denominators.

For sums of k unit fractions we adopt the notation of [9] and define $f_k(m,n)$ to be the number of solutions $(a_1, a_2, \ldots, a_k) \in \mathbb{N}^k$ of equation (2.2) with $a_1 \leq a_2 \leq \ldots \leq a_k$, i.e.

$$f_k(m,n) = \left| \left\{ (a_1, a_2, \ldots, a_k) \in \mathbb{N}^k : \frac{m}{n} = \frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_k}, a_1 \leq a_2 \ldots \leq a_k \right\} \right|.$$

Concerning equation (2.1) with $m = 4$ the results of Elsholtz and Tao [25] show that the number of solutions $f_3(4,n)$ is related to some divisor questions and is on average a power of $\log n$ (at least when n is prime). It even seems possible that for fixed $m \in \mathbb{N}$ and any $\epsilon > 0$ the number of representations of $\frac{m}{n}$ as a sum of k unit fractions is bounded by $\mathcal{O}_{k,\epsilon}(n^\epsilon)$. More details on this are informally and heuristically discussed in Section 2.3. For general m and n the best known upper bound on the number of solutions of (2.1) is due to Browning and Elsholtz [9, Theorem 2] who proved an upper bound of order $\mathcal{O}_\epsilon(n^\epsilon \left(\frac{n}{m}\right)^{2/3})$. In the case of the Erdős-Straus equation with $n = p$ prime Elsholtz and Tao [25, Proposition 1.7] have improved this bound to $\mathcal{O}_\epsilon(p^{3/5+\epsilon})$. It is known that this type of question is easier to study, when the denominator is prime.

Our main result will be the following theorem which provides an upper bound on the number of solutions of equation (2.1).

Theorem 2.2. *For any $m, n \in \mathbb{N}$ and any $\epsilon > 0$ there are at most $\mathcal{O}_\epsilon \left(n^\epsilon \left(\frac{n^3}{m^2} \right)^{1/5} \right)$ solutions of the equation*

$$\frac{m}{n} = \frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3}$$

in positive integers a_1, a_2 and a_3.

Note that this improves upon the bound of Browning and Elsholtz in the range $m \ll n^{1/4}$. As a corollary we get that the Elsholtz-Tao bound for the number of solutions of the Erdős-Straus equation is true for arbitrary denominators $n \in \mathbb{N}$.

Corollary 2.3. *The Erdős-Straus equation*

$$\frac{4}{n} = \frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3}$$

has at most $\mathcal{O}_\epsilon(n^{3/5+\epsilon})$ *solutions in positive integers* a_1, a_2 *and* a_3.

We also prove the following algorithmic version of Theorem 2.2 with a matching upper bound for the expected running time[1].

Corollary 2.4. *There exists an algorithm with an expected running time of order*

$$\mathcal{O}_\epsilon\left(n^\epsilon\left(\frac{n^3}{m^2}\right)^{1/5}\right),$$

for any $\epsilon > 0$*, which lists all representations of the rational number* $\frac{m}{n}$ *as a sum of three unit fractions. Furthermore, all representations of* $\frac{m}{n}$ *as a sum of* $k > 3$ *unit fractions may be found in expected time* $\mathcal{O}_{\epsilon,k}\left(n^{2^{k-3}(8/5+\epsilon)-1}\right)$*, for any* $\epsilon > 0$.

For sums of k unit fractions we will prove the following result.

Theorem 2.5. *We have*

$$f_4(m,n) \ll_\epsilon n^\epsilon\left(\frac{n^{4/3}}{m^{2/3}} + \frac{n^{28/17}}{m^{8/5}}\right)$$

and for any $k \geq 5$

$$f_k(m,n) \ll_\epsilon (kn)^\epsilon\left(\frac{k^{4/3}n^2}{m}\right)^{28/17 \cdot 2^{k-5}}.$$

Keeping in mind that $\frac{28}{17} = 1.64705\ldots$, Theorem 2.5 may be compared with the following bounds from [9, Theorem 3]:

$$f_4(m,n) \ll_\epsilon n^\epsilon\left(\frac{n^{4/3}}{m^{2/3}} + \left(\frac{n}{m}\right)^{5/3}\right),$$

$$f_k(m,n) \ll_\epsilon (kn)^\epsilon\left(\frac{k^{4/3}n^2}{m}\right)^{5/3 \cdot 2^{k-5}}, \text{ for } k \geq 5.$$

A well studied special case of Theorem 2.5 concerns representations of 1 as a sum of k unit fractions. Browning and Elsholtz [9] mention several related problems which are studied in the literature and can be improved using better upper bounds on $f_k(m,n)$. We summarize these results in the following corollary.

Corollary 2.6. *1. For any* $\epsilon > 0$ *we have that*

$$f_k(1,1) \ll_\epsilon k^{7/51 \cdot 2^{k-1}+\epsilon}.$$

[1] For a definition of expected running time see the proof of this corollary at the end of Section 2.5.

2. *Let c_0 be as in Definition 2.1. Then for $\epsilon > 0$ and $k \geq k(\epsilon)$, we have*

$$f_k(1,1) < c_0^{(7/17+\epsilon)2^{k-1}}.$$

3. *For $\epsilon > 0$ and $k \geq k(\epsilon)$ the number $S(k)$ of positive integer solutions of the equation*

$$1 = \sum_{i=1}^{k} \frac{1}{a_i} + \frac{1}{\prod_{i=1}^{k} a_i}$$

is bounded from above by $c_0^{(7/17+\epsilon)2^k}$.

Proof. The first assertion is an immediate consequence of Theorem 2.5. For the proof of the second statement we refer the reader to the proof of Theorem 4 in [9]. The only change necessary is plugging in the bound from Theorem 2.5 instead of [9, Theorem 3] for the last 5 lines of the proof which amounts to just exchanging one exponent. The last statement follows from the first one and the observation that $S(k) \leq f_{k+1}(1,1)$. $\square$

We note that the number of solutions of the equation $1 = \sum_{1=1}^{k} \frac{1}{a_i} + \frac{1}{\prod_{i=1}^{k} a_i}$ has applications to problems considered in [8].

Finally we deal with lower bounds. In [25, Theorem 1.8] it is shown that we have

$$f_3(4,n) \geq \exp\left((\log 3 + o(1))\frac{\log n}{\log\log n}\right)$$

for infinitely many $n \in \mathbb{N}$ and that

$$f_3(4,n) \geq \exp\left(\left(\frac{\log 3}{2} + o(1)\right)\log\log n\right)$$

for all integers n in a subset of the positive integers with density 1. The following theorem gives an improvement of these bounds which also give a limitation on improving the upper bounds for the number of solution of the Erdős-Straus equation and in the general case. For comparison we note that $\log 3 = 1.09861\ldots$, $\frac{\log 3}{2} = 0.54930\ldots$ and $\log 6 = 1.79175\ldots$.

Theorem 2.7. *For given $m \in \mathbb{N}$ there are infinitely many $n \in \mathbb{N}$ such that*

$$f_3(m,n) \geq \exp\left((\log 6 + o_m(1))\frac{\log n}{\log\log n}\right).$$

Furthermore, for given $m \in \mathbb{N}$, there exists a subset $\mathcal{M}_1$ of the positive integers, which *has density one, such that for any* $n \in \mathcal{M}_1$

$$\begin{aligned} f_3(m,n) &\geq \exp\left((\log 3 + o_m(1)) \log\log n\right) \cdot \log\log n \\ &\gg (\log n)^{\log 3 + o_m(1)}. \end{aligned}$$

For the special case $m = 4$ *and for integers* n *in a set* $\mathcal{M}_2 \subset \mathbb{N}$ *with density one, the last bound may be improved to*

$$f_3(4,n) \geq \exp\left((\log 6 + o(1)) \log\log n\right).$$

Remark 2.8. *Previous proofs of lower bounds of similar type as the ones in Theorem 2.7 constructed solutions from factorizations of* n*. We get our improvement from additionally taking into account factorizations of a lot of shifts of* n*. Hence our proof also shows that there are many values* a_1 *admitting many pairs* (a_2, a_3)*. Here, depending on which of the three lower bounds in Theorem 2.7 we consider, 'many' may either mean* $\exp\left((C + o_m(1))\frac{\log n}{\log\log n}\right)$ *or* $\exp\left((\tilde{C} + o_m(1))\log\log n\right)$*, for suitable positive constants* C *and* $\tilde{C}$ *.*

We may ask if a lower bound on $f_3(m,n)$ of the first type in Theorem 2.7 does not only hold for infinitely many positive integers n but also for infinitely many prime denominators p. In [25] there was no lower bound of this type, but it was proved that $f_3(4,p) \gg (\log p)^{0.549}$ for almost all primes. We note that this result implies, using Dirichlet's theorem on primes, the following corollary.

Corollary 2.9. *For every reduced residue class* $e \bmod f$*, i.e.* $\gcd(e,f) = 1$*, there are infinitely many primes* p *such that* $f_3(4,p) \gg (\log p)^{0.549}$*, and* $p \equiv e \bmod f$*.*

Here we improve this corollary considerably.

Theorem 2.10. *For every* $m \in \mathbb{N}$ *and every reduced residue class* $e \bmod f$ *there are infinitely many primes* $p \equiv e \bmod f$ *such that*

$$f_3(m,p) \gg_{f,m} \exp\left(\left(\frac{5\log 2}{12\,\mathrm{lcm}(m,f)} + o_{f,m}(1)\right)\frac{\log p}{\log\log p}\right).$$

Here $o_{f,m}(1)$ *denotes a quantity depending on* f *and* m *which goes to zero as* p *tends to infinity.*

2.2. Notation

As usual $\mathbb{N}$ denotes the set of positive integers and $\mathbb{P}$ the set of primes in $\mathbb{N}$. We denote the greatest common divisor and the least common multiple of n elements $a_i \in \mathbb{N}$ by $\gcd(a_1, a_2, \ldots, a_n)$ and $\operatorname{lcm}(a_1, a_2, \ldots, a_n)$ or $(a_1, a_2, \ldots, a_n)$ and $[a_1, a_2, \ldots, a_n]$ for short. For integers $d, n \in \mathbb{N}$ we write $d|n$ if d divides n. We use the symbols $\mathcal{O}$, o, $\ll$ and $\gg$ within the contexts of the well known Landau and Vinogradov notations where dependence of the implied constant on certain variables is indicated by a subscript. For any prime $p \in \mathbb{P}$ we define the function $\nu_p : \mathbb{N} \to \mathbb{N} \cup \{0\}$ to be the p-adic valuation, i.e. $\nu_p(n) = a$ if and only if p^a is the highest power of p dividing n. By $\tau(n)$ and $\omega(n)$, as usual, we denote the number of divisors and the number of distinct prime divisors of n. By $\tau(n, m)$, we denote the number of divisors of n coprime to m and $\tau(n, k, m)$, $\omega(n, k, m)$ denote the number of divisors (resp. distinct prime divisors) of n in the residue class $k \bmod m$, where $(k, m) = 1$. Finally, for two coprime integers a and b we denote by $\operatorname{ord}_a(b)$ the least positive integer l, such that $b^l \equiv 1 \bmod a$.

2.3. Heuristics on $f_k(m,n)$

We now informally discuss why $f_3(m, n) = \mathcal{O}_\epsilon(n^\epsilon)$ can be expected. In fact, as far as we are aware, this was first observed by Roger Heath-Brown (private communication with the first author in 1994). Let us first recall (see e.g. [71, p. 201: Theorem 3]) that a fraction $\frac{m}{n}$ with $\gcd(m, n) = 1$ is a sum of two unit fractions $\frac{1}{a_1} + \frac{1}{a_2}$ if and only if there exist two distinct, positive and coprime divisors d_1 and d_2 of n such that $d_1 + d_2 \equiv 0 \bmod m$. We may deduce an upper bound of $\mathcal{O}_\epsilon(n^\epsilon)$ for the number of representations of $\frac{m}{n}$ as a sum of two unit fractions. Indeed from

$$\frac{m}{n} = \frac{1}{a_1} + \frac{1}{a_2}, \tag{2.3}$$

by setting $d = (a_1, a_2)$ and $a_i' = \frac{a_i}{d}$ for $i \in \{1, 2\}$, we see that

$$m a_1' a_2' d = n(a_1' + a_2').$$

This implies that a_1', a_2' are divisors of n, d divides $n(a_1' + a_2') < 2n^2$ and any solution (a_1, a_2) of (2.3) uniquely corresponds to a triple (a_1', a_2', d). The number $\sum_{a_1', a_2' | n} \tau(n(a_1' + a_2'))$ of such triples is bounded by $\mathcal{O}_\epsilon(n^\epsilon)$ (see Lemma 2.12 below).

Studying $\frac{m}{n} = \frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3}$ with $a_1 \le a_2 \le a_3$ one observes that

$$\frac{1}{a_1} < \frac{m}{n} \le \frac{3}{a_1}$$

from which $\frac{n}{m} < a_1 \le \frac{3n}{m}$ follows. In view of

$$\frac{m}{n} - \frac{1}{a_1} = \frac{ma_1 - n}{na_1} = \frac{1}{a_2} + \frac{1}{a_3} \tag{2.4}$$

there are at most $\mathcal{O}\left(\frac{n}{m}\right)$ choices for a_1, and for given a_1 there are at most $d(na_1) = \mathcal{O}_\epsilon(n^\epsilon)$ divisors of na_1. This shows that $f_3(m,n) = \mathcal{O}_\epsilon\left(\frac{n^{1+\epsilon}}{m}\right)$ is a trivial upper bound. The real question is for how many values of a_1 there can be at least one solution. For increasing a_1, even if na_1 contains many divisors, the congruence $d_1 + d_2 \equiv 0 \bmod ma_1 - n$ should become, on average, more difficult to satisfy if $ma_1 - n \gg n^\epsilon$. Therefore, we expect that the number of a_1 contributing at least one solution is $\mathcal{O}_\epsilon(n^\epsilon)$, so that $f_3(m,n) = \mathcal{O}_\epsilon(n^{2\epsilon})$. Moreover, equation (2.4) implies that for any given a_1, the number of solutions is about $\tilde{d}(m,n,a_1)$. Here $\tilde{d}(m,n,a_1)$ counts the number of pairs of coprime divisors d_1, d_2 of na_1, with $d_1 + d_2 \equiv 0 \bmod ma_1 - n$. Therefore, $f_3(m,n)$ should be approximately $\sum_{a_1} \tilde{d}(m,n,a_1)$.

Similarly a completely trivial upper bound on $f_4(m,n)$ is as follows. With $a_1 \le a_2 \le a_3 \le a_4$ it follows that $\frac{n}{m} < a_1 \le \frac{4n}{m}$ and hence

$$\frac{ma_1 - n}{na_1} = \frac{m}{n} - \frac{1}{a_1} = \frac{1}{a_2} + \frac{1}{a_3} + \frac{1}{a_4} \le \frac{3}{a_2}.$$

From those bounds we easily deduce that $a_2 \le \frac{12n^2}{m}$. With

$$\frac{m}{n} - \frac{1}{a_1} - \frac{1}{a_2} = \frac{ma_1a_2 - na_2 - na_1}{na_1a_2} = \frac{1}{a_3} + \frac{1}{a_4},$$

with similar arguments as above, we deduce that $f_4(m,n) = \mathcal{O}_\epsilon\left(\frac{n^{3+\epsilon}}{m^2}\right)$. For fixed m the fact that our bound on $f_4(m,n)$ in Theorem 2.5 above is better than $\mathcal{O}(n^2)$ shows that, for most pairs (a_1, a_2) and moreover, for most choices of $a_2 \in \left[\frac{n}{m}, \frac{12n^2}{m}\right]$ there is no solution of $\frac{m}{n} = \frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3} + \frac{1}{a_4}$. Here again, as soon as $ma_1a_2 - na_2 - na_1 \gg n^\epsilon$ one should not expect to have two divisors d_1, d_2 of na_1a_2 such that $d_1 + d_2 \equiv 0 \bmod ma_1a_2 - na_2 - na_1$. From this reasoning, also $f_k(m,n) = \mathcal{O}_{\epsilon,k}(n^\epsilon)$, for $k \ge 4$ seems to us a reasonable expectation.

The papers [9] and [25] studied parametric solutions of the diophantine equation (2.1). The reason why the result in [25] is superior in the case of n being a prime is that here a full parametric solution (e.g. [67]) is much easier to work with. However, in this manuscript we develop parametric solutions of (2.1) and (2.2) from scratch. Some simplified version of this has been used in [19] and [25, Section 11], but there the focus was to generate solutions with many parameters. Here we need to do kind of the opposite, namely to show that *every* solution comes from a number of parametric families.

The method we introduce should theoretically work for any diophantine equation as it expresses a k-tuple of integers in a standard form. In practice it might work favorably if there is some inhomogeneous part as in

$$n = a_1 a_2 a_3 - a_1 - a_2.$$

For prime values of n in equation (2.1) there are several discussions of parametric solutions in the literature, e.g. by Rosati [67] and Aigner [1], see also Mordell's book [62, Chapter 30]. For composite values n there is no satisfactory treatment in the literature, and Section 2.5 below may be the most detailed study to date.

2.4. Patterns and relative greatest common divisors

Consider a solution $(a_1, a_2, \ldots, a_k) \in \mathbb{N}^k$ with $a_1 \le a_2 \le \ldots \le a_k$ of equation (2.2) and set $n_i = (a_i, n)$, $a_i = n_i t_i$ for $i \in \{1, 2, \ldots, k\}$. We can thus rewrite equation (2.2) as

$$\frac{m}{n} = \sum_{i=1}^{k} \frac{1}{n_i t_i}. \tag{2.5}$$

Later, when working on upper bounds for the number of solutions of equation (2.5) for $k \in \{3, 4\}$, we will fix a choice of $(n_1, n_2, \ldots, n_k) \in \mathbb{N}^k$. For given $m, n \in \mathbb{N}$ we call such a choice the *pattern* of a solution of this equation. Note that for solutions corresponding to a given pattern $(n_1, n_2, \ldots, n_k)$ we have that $\left(\frac{n}{n_i}, t_i\right) = 1$ for all $i \in \{1, 2, \ldots, k\}$. As $n_i | n$ the number of distinct patterns is $\mathcal{O}_k(n^\epsilon)$ only.

Also, when dealing with equations of type (2.5) for $k \in \{3, 4\}$ we will make heavy use of the concept of relative greatest common divisors as described by Elsholtz in [18] (for some ad hoc definition see also [19]). Relative greatest common divisors are a useful tool when studying divisibility relations among the t_i in (2.5).

Let $I = \{1, 2, \ldots, k\}$ be the index set. Then we define the relative greatest common divisors of the positive integers $t_1, t_2, \ldots, t_k$ recursively as follows:

$$x_I = \gcd(t_1, t_2, \ldots, t_k)$$

and for any $\{i_1, i_2, \dots i_{|J|}\} = J \subseteq I,\ J \neq \emptyset$ we set

$$x_J = \frac{\gcd(t_{i_1}, t_{i_2}, \dots, t_{i_{|J|}})}{\prod_{\substack{J' \subseteq I \\ J \subsetneq J'}} x_{J'}}.$$

For $k \in \{3, 4\}$ we will later identify the elements x_J with $J \subseteq I$ with the elements x_i, x_{ij} and x_{ijk} where $\{i, j, k\} = \{1, 2, 3\}$ in the case when $k = 3$ and with the elements x_i, x_{ij}, x_{ijk} and x_{ijkl} with $\{i, j, k, l\} = \{1, 2, 3, 4\}$ when $k = 4$. With the relative greatest common divisors defined as above we have that

$$t_i = \prod_{\substack{J \subseteq I \\ i \in J}} x_J.$$

A further very useful property of relative greatest common divisors is that $(x_J, x_K) = 1$ if $J \nsubseteq K$ and $K \nsubseteq J$. We prove this property as the following lemma (see also [18, p. 2]).

Lemma 2.11. *Let $t_1, t_2, \dots, t_k \in \mathbb{N}$, $J, K \subseteq \{1, 2, \dots, k\}$, $J, K \neq \emptyset$ and define the corresponding relative greatest common divisors x_J and x_K as above. If $J \nsubseteq K$ and $K \nsubseteq J$ then $(x_J, x_K) = 1$.*

Proof. By assumption $J \nsubseteq K$ and $K \nsubseteq J$ and thus we have that $J \subsetneq J \cup K$ and $K \subsetneq J \cup K$. We suppose that $d = (x_J, x_K) > 1$ and choose an arbitrary prime divisor $p|d$. Set $L = J \cup K$, $J = \{j_1, j_2, \dots, j_{|J|}\}$, $K = \{k_1, k_2, \dots, k_{|K|}\}$, $L = \{l_1, l_2, \dots, l_{|L|}\}$ and write

$$x_J = \frac{(t_{j_1}, t_{j_2}, \dots, t_{j_{|J|}})}{\left(\prod_{\substack{J' \subseteq I \\ J \subsetneq J' \\ L \nsubseteq J'}} x_{J'}\right) \cdot x_L \cdot \left(\prod_{\substack{J' \subseteq I \\ L \subsetneq J'}} x_{J'}\right)},$$

$$x_K = \frac{(t_{k_1}, t_{k_2}, \dots, t_{k_{|K|}})}{\left(\prod_{\substack{K' \subseteq I \\ K \subsetneq K' \\ L \nsubseteq K'}} x_{K'}\right) \cdot x_L \cdot \left(\prod_{\substack{K' \subseteq I \\ L \subsetneq K'}} x_{K'}\right)}.$$

With $x_L = \frac{(t_{l_1}, t_{l_2}, \ldots, t_{l_{|L|}})}{\prod_{\substack{L' \subseteq I \\ L \subsetneq L'}} x_{L'}}$ this simplifies to

$$x_J = \frac{(t_{j_1}, t_{j_2}, \ldots, t_{j_{|J|}})}{\left(\prod_{\substack{J' \subseteq I \\ J \subsetneq J' \\ L \not\subseteq J'}} x_{J'}\right) \cdot (t_{l_1}, t_{l_2}, \ldots, t_{l_{|L|}})}, \; x_K = \frac{(t_{k_1}, t_{k_2}, \ldots, t_{k_{|K|}})}{\left(\prod_{\substack{K' \subseteq I \\ K \subsetneq K' \\ L \not\subseteq K'}} x_{K'}\right) \cdot (t_{l_1}, t_{l_2}, \ldots, t_{l_{|L|}})}. \quad (2.6)$$

Let p^α be the highest power of p dividing the greatest common divisor of the terms

$$(t_{j_1}, t_{j_2}, \ldots, t_{j_{|J|}}) \text{ and } (t_{k_1}, t_{k_1}, \ldots, t_{k_{|K|}}).$$

Thus p^α is also the highest power of p such that

$$p^\alpha | ((t_{j_1}, t_{j_2}, \ldots, t_{j_{|J|}}), (t_{k_1}, t_{k_1}, \ldots, t_{k_{|K|}})) = (t_{l_1}, t_{l_2}, \ldots, t_{l_{|L|}}).$$

By definition of the greatest common divisor, without loss of generality we may suppose that $\nu_p((t_{j_1}, t_{j_2}, \ldots, t_{j_{|J|}})) = \alpha$. From equation (2.6) we finally see that $\nu_p(x_J) = 0$, a contradiction to $p|d$. $\square$

Relative greatest common divisors may be nicely visualized via Venn diagrams (especially when $k \leq 3$). We identify a positive integers with the multiset of its prime divisors, i.e. each prime p dividing n occurs with multiplicity $\nu_p(n)$ in the multiset. Given the Venn diagram of the multisets corresponding to the integers $t_1, \ldots, t_k$, each area of intersection in the diagram uniquely corresponds to a relative greatest common divisor x_J, $J \subseteq \{1, \ldots, k\}$. Figure 2.1 shows the situation for relative greatest common divisors of three positive integers t_1, t_2 and t_3.

As mentioned in the beginning of this section relative greatest common divisors were systematically described in [18]. Nonetheless concepts of a similar type date back at least as far as Dedekind [15] who called the relative greatest common divisors of the integers $t_1, \ldots, t_k$ the cores (Kerne) of the system $(t_1, \ldots, t_k)$. Dedekind described the construction of these cores explicitly for systems with three and four elements and developed some theory to describe the cores of systems with more than four elements.

Decompositions similar to relative greatest common divisors also occur when we look for generalizations of the formula

$$[t_1, t_2] = \frac{t_1 t_2}{(t_1, t_2)}, \quad (2.7)$$

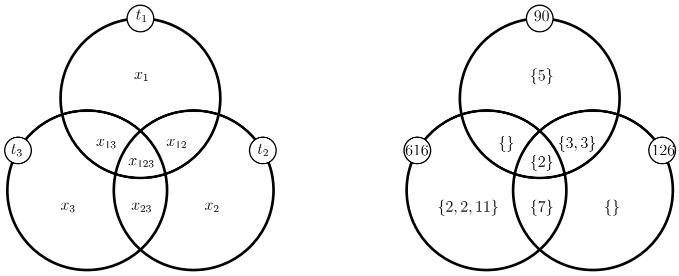

Figure 2.1.: A visualization of relative greatest common divisors using Venn diagrams. On the left hand side one sees the general case of three positive integers t_1, t_2 and t_3 and on the right hand side the situation when $t_1 = 90$, $t_2 = 126$ and $t_3 = 616$. Empty sets correspond to empty products and we set the corresponding relative greatest common divisor to 1.

where $[t_1, t_2]$ denotes the least common multiple of the integers t_1 and t_2. A generalization of formula (2.7) to least common multiples and greatest common divisors of k integers $t_1, \dots, t_k$ was found by V.-A. Lebesgue [54, p. 350], who proved that

$$[t_1, t_2, \dots, t_k] = \frac{\prod_{\substack{1 \le i \le k \\ i \text{ odd}}} G_i}{\prod_{\substack{1 \le j \le k \\ j \text{ even}}} G_j},$$

where the variables G_i denote the product of the greatest common divisors of all choices of subsets of i integers in the set $\{t_1, t_2, \dots, t_k\}$.

2.5. Sums of three unit fractions

In this section we deal with equation (2.5) for $k = 3$, i.e. with equations of the form

$$\frac{m}{n} = \frac{1}{n_1 t_1} + \frac{1}{n_2 t_2} + \frac{1}{n_3 t_3}, \tag{2.8}$$

where $n_1 t_1 \le n_2 t_2 \le n_3 t_3$, $n_i | n$ and $\left(\frac{n}{n_i}, t_i\right) = 1$ for $i \in \{1, 2, 3\}$. In the following we use the concept of relative greatest common divisors introduced in the previous section to get a suitable parametrisation of the solutions of (2.8) corresponding to a fixed pattern $(n_1, n_2, n_3) \in \mathbb{N}^3$.

Writing the variables t_i in terms of relative greatest common divisors, equation (2.8) takes the form

$$\frac{m}{n} = \frac{1}{n_1 x_1 x_{12} x_{13} x_{123}} + \frac{1}{n_2 x_2 x_{12} x_{23} x_{123}} + \frac{1}{n_3 x_3 x_{13} x_{23} x_{123}} \tag{2.9}$$

and multiplying out yields

$$mx_1x_2x_3x_{12}x_{13}x_{23}x_{123} = \frac{n}{n_1}x_2x_3x_{23} + \frac{n}{n_2}x_1x_3x_{13} + \frac{n}{n_3}x_1x_2x_{12}. \tag{2.10}$$

A first thing we observe is that we have $x_i = 1$ for all $i \in \{1,2,3\}$. This follows from Lemma 2.11 and equation (2.10) together with the fact that $x_i|\frac{n}{n_i}$ is possible only if $x_i = 1$ by definition of n_i. We thus can work with the following simplified version of equation (2.10)

$$mx_{12}x_{13}x_{23}x_{123} = \frac{n}{n_1}x_{23} + \frac{n}{n_2}x_{13} + \frac{n}{n_3}x_{12}. \tag{2.11}$$

Next we introduce the parameters d_{ij} which are defined as $d_{ij} = \left(\frac{n}{n_i}, \frac{n}{n_j}\right)$. Again we have that $(x_{ij}, d_{ij}) = 1$ by definition of the n_i and we note that for given m, n and a fixed pattern (n_1, n_2, n_3) also the parameters d_{ij} are fixed.

In what follows we apply methods developed by Elsholtz and Tao [25, Sections 2 and 3]. The strategy is to derive a system of equations from (2.11) and to make use of divisor relations therein. With the observation of coprimality of d_{ij} and x_{ij}, and using divisibility relations implied by equation (2.11) we may define the following three positive integers

$$w = \frac{\frac{n}{n_1d_{13}}x_{23} + \frac{n}{n_3d_{13}}x_{12}}{x_{13}}, \; y = \frac{\frac{n}{n_1d_{12}}x_{23} + \frac{n}{n_2d_{12}}x_{13}}{x_{12}} \text{ and } z = \frac{\frac{n}{n_2d_{23}}x_{13} + \frac{n}{n_3d_{23}}x_{12}}{x_{23}}.$$

Later we make use of the product of w and z which is given by

$$\begin{aligned} wz &= \frac{n}{n_1d_{13}}\frac{n}{n_2d_{23}} + \frac{x_{12}}{x_{13}x_{23}}\left(\frac{n^2}{n_1n_3d_{13}d_{23}}x_{23} + \frac{n^2}{n_2n_3d_{13}d_{23}}x_{13} + \frac{n^2}{n_3^2d_{13}d_{23}}x_{12}\right) \\ &= \frac{n}{n_1d_{13}}\frac{n}{n_2d_{23}} + \frac{nx_{12}}{n_3d_{13}d_{23}x_{13}x_{23}}\left(\frac{n}{n_1}x_{23} + \frac{n}{n_2}x_{13} + \frac{n}{n_3}x_{12}\right) \\ &= \frac{n}{n_1d_{13}}\frac{n}{n_2d_{23}} + \frac{nm}{n_3d_{13}d_{23}}x_{12}^2x_{123}, \end{aligned}$$

where we used equation (2.11) to get the last equality. We collect the equations just derived in the following list

$$mx_{12}x_{13}x_{23}x_{123} = \frac{n}{n_1}x_{23} + \frac{n}{n_2}x_{13} + \frac{n}{n_3}x_{12} \tag{2.12}$$

$$yx_{12} = \frac{n}{n_1d_{12}}x_{23} + \frac{n}{n_2d_{12}}x_{13} \tag{2.13}$$

$$zx_{23} = \frac{n}{n_2d_{23}}x_{13} + \frac{n}{n_3d_{23}}x_{12} \tag{2.14}$$

$$mx_{13}x_{23}x_{123} = d_{12}y + \frac{n}{n_3} \tag{2.15}$$

$$mx_{12}x_{13}x_{123} = d_{23}z + \frac{n}{n_1} \tag{2.16}$$

$$wz = \frac{n}{n_1 d_{13}} \frac{n}{n_2 d_{23}} + \frac{nm}{n_3 d_{13} d_{23}} x_{12}^2 x_{123}. \tag{2.17}$$

For proving Theorem 2.2 the classical divisor bound will play a crucial role. We will use it in the following form (see [47, Theorem 315]).

Lemma 2.12 (Divisor bound)**.** *Let $d(n) : \mathbb{N} \to \mathbb{N}$ be the divisor function, i.e. $d(n) = \sum_{d|n} 1$. Then for every $\epsilon > 0$, we have*

$$d(n) \ll_\epsilon n^\epsilon.$$

We now have all the tools we need to prove Theorem 2.2.

Proof of Theorem 2.2. Consider a solution of equation (2.8) for a fixed pattern (n_1, n_2, n_3). By assumption we have $n_1 t_1 \leq n_2 t_2 \leq n_3 t_3$ and using the parametrization of the t_i we introduced in equation (2.9) this implies

$$x_{13} \leq \frac{n_2}{n_1} x_{23} \text{ and } x_{12} \leq \frac{n_3}{n_2} x_{13}.$$

Using these inequalities in equations (2.13) and (2.14) yields

$$yx_{12} \leq 2\frac{n}{n_1 d_{12}} x_{23} \text{ and } zx_{23} \leq 2\frac{n}{n_2 d_{23}} x_{13}.$$

Dividing by x_{23} and x_{13} respectively and multiplying the last two inequalities we arrive at

$$\frac{yx_{12}}{x_{23}} \frac{zx_{23}}{x_{13}} \leq 4\frac{n^2}{n_1 n_2 d_{12} d_{23}}.$$

We now intend to obtain a lower bound for $n_1 n_2 d_{12} d_{23}$. Let $n = \prod_{p \in \mathbb{P}} p^{\nu_p(n)}$ be the prime factorization of n. Then $n_1 = \prod_{p \in \mathbb{P}} p^{\nu_p(n_1)}$ and $n_2 = \prod_{p \in \mathbb{P}} p^{\nu_p(n_2)}$ where $0 \leq \nu_p(n_1), \nu_p(n_2) \leq \nu_p(n)$ for all $p \in \mathbb{P}$. Since

$$d_{12} = \left(\frac{n}{n_1}, \frac{n}{n_2}\right) = \prod_{p \in \mathbb{P}} p^{\nu_p(n) - \max(\nu_p(n_1), \nu_p(n_2))}$$

we have

$$n_1 n_2 d_{12} = \prod_{p \in \mathbb{P}} p^{\nu_p(n_1) + \nu_p(n_2) + \nu_p(n) - \max(\nu_p(n_1), \nu_p(n_2))}$$

$$\geq \prod_{p\in\mathbb{P}} p^{\nu_p(n_1)+\nu_p(n_2)+\nu_p(n)-\nu_p(n_1)-\nu_p(n_2)} = n.$$

This shows that $n_1 n_2 d_{12} d_{23} \geq n$ and thus

$$\frac{yx_{12}}{x_{23}}\frac{zx_{23}}{x_{13}} \ll n.$$

By assumption we have that $n_1 t_1$ is the smallest denominator in equation (2.8). This implies that

$$\frac{m}{n} \leq \frac{3}{n_1 t_1} \text{ and thus } t_1 \leq \frac{3n}{mn_1} \ll \frac{n}{m}.$$

The bound in Theorem 2.2 can finally be derived from the following inequality

$$y \cdot z \cdot x_{12}x_{13} \cdot (x_{12}x_{123})^2 = \frac{yx_{12}}{x_{23}}\frac{zx_{23}}{x_{13}}(x_{12}x_{13}x_{123})^2 \ll \frac{n^3}{m^2}. \tag{2.18}$$

This implies that at least one of the factors y, z, $x_{12}x_{13}$ and $x_{12}x_{123}$ is bounded by

$$\mathcal{O}\left(\left(\frac{n^3}{m^2}\right)^{1/5}\right).$$

If this is the case for y then by Lemma 2.12 and equation (2.15) we have at most $\mathcal{O}_\epsilon(n^\epsilon)$ choices for the parameters x_{13}, x_{23} and x_{123} for every choice of y. The parameter x_{12} is then uniquely determined by (2.12).

Similarly, if z is the bounded parameter use Lemma 2.12 and equation (2.16) to see that there are at most $\mathcal{O}_\epsilon(n^\epsilon)$ choices for the parameters x_{12}, x_{13} and x_{123} for every choice of z. Again the remaining parameter x_{23} is uniquely determined by (2.12).

Suppose that $x_{12}x_{13} \ll \left(\frac{n^3}{m^2}\right)^{1/5}$. By Lemma 2.12 for every fixed choice of $x_{12}x_{13}$ we may choose the factors x_{12} and x_{13} in at most $\mathcal{O}_\epsilon(n^\epsilon)$ ways. For each of those choices Lemma 2.12 and equation (2.14) imply that there are at most $\mathcal{O}_\epsilon(n^\epsilon)$ choices for the parameter x_{23}. As before the remaining parameter x_{123} is then fixed by (2.12).

Finally we need to consider the case when $x_{12}x_{123}$ is the bounded factor. As in the previous case for any fixed choice of $x_{12}x_{123}$ we have at most $\mathcal{O}_\epsilon(n^\epsilon)$ choices for the factors x_{12} and x_{123}. Since equation (2.8) has no solutions for $m > 3n$ we have that $m \ll n$ and using equation (2.17) we see that for any fixed choice of x_{12} and x_{123} we have at most $\mathcal{O}_\epsilon(n^\epsilon)$ choices for the parameters w and z. With z, x_{12} and x_{123} fixed, x_{13} is uniquely determined by (2.16). The last parameter x_{23} is again uniquely determined by (2.12).

In any case we have a bounded number of applications of the divisor bound from

Lemma 2.12, say it was applied at most l times to integers of size at most $\mathcal{O}(n^c)$, for a fixed constant c [2]. Setting $\tilde{\epsilon} = cl\epsilon$ we hence have at most $\mathcal{O}_{\tilde{\epsilon}}\big(n^{\tilde{\epsilon}}\big(\frac{n^3}{m^2}\big)^{1/5}\big)$ choices for the parameters x_{12}, x_{13}, x_{23} and x_{123} which uniquely determine a solution of (2.8) if n_1, n_2 and n_3 are fixed. Note that this bound is independent of the concrete choice of the parameters n_i and again by Lemma 2.12 we have at most $\mathcal{O}_\epsilon(n^{3\epsilon})$ choices for the pattern (n_1, n_2, n_3). Theorem 2.2 now follows by redefining the choice of ϵ. □

Finally we prove Corollary 2.4.

Proof of Corollary 2.4. The proof of Theorem 2.2 suggests an algorithm for computing all decompositions of a rational number $\frac{m}{n}$ as a sum of three unit fractions. The running time of this algorithm depends on the quality of algorithms used for integer factorization. In [56] a probabilistic algorithm is analyzed which finds all prime factors of a given integer in expected running time $\exp((1 + o(1))\sqrt{\log n \log\log n})$ for $n \to \infty$, which is clearly $\mathcal{O}_\epsilon(n^\epsilon)$. Lenstra and Pomerance [56, Section 12] point out, that here the term probabilistic means that the algorithm is allowed to call a random number generator which outputs 0 or 1 each with probability $\frac{1}{2}$. The term expected running time refers to averaging over the output of the random number generator only and not over the input n. Hence the expected running time is also valid for each individual n.

As a consequence, using an algorithm of this type, all decompositions of $\frac{m}{n}$ as a sum of three unit fractions can be found by carrying out the following steps. Factorize the integer n and compute all possible patterns (n_1, n_2, n_3). For any of these $\mathcal{O}_\epsilon(n^\epsilon)$ patterns it follows from the calculations in the proof of Theorem 2.2, that the implied constant in inequality (2.18) may be chosen as $C := \big(\frac{36}{n_1^2 d_{23}}\big)$. For all choices of integers y, z, $x_{12}x_{13}$ and $x_{12}x_{123} \in \big[1, C^{1/5}\big(\frac{n^3}{m^2}\big)^{1/5}\big]$ we determine the integers x_{12}, x_{13}, x_{23} and x_{123} via factoring $x_{12}x_{13}$, $x_{12}x_{123}$ and a small number of integers mentioned in formulae (2.12)-(2.17). All in all this leads to an algorithm of expected running time $\mathcal{O}_\epsilon\big(n^\epsilon\big(\frac{n^3}{m^2}\big)^{1/5}\big)$.

As for representations of the form

$$\frac{m}{n} = \sum_{i=1}^{k} \frac{1}{a_i} \tag{2.19}$$

with $k > 3$ we enumerate all possible choices for the denominators a_i, $1 \le i \le k-3$, and apply our algorithm for finding representations as sum of three unit fractions to

[2] It is easy to see that the largest denominator n_3t_3 in equation (2.8) is bounded by $\mathcal{O}(n^4)$ (see also the proof of Corollary 2.4 below). The same bound hence applies to all the parameters x_i, x_{ij} and x_{123}, $\{i, j, k\} = \{1, 2, 3\}$. Since all integers we need to factor are either products of two of these relative greatest common divisors, or appear on one side of the equations (2.12) – (2.17), together with the fact that $m \le 3n$, this implies the existence of the constant c.

determine all choices for the remaining three denominators, i.e. we solve

$$\frac{m}{n} - \sum_{i=1}^{k-3} \frac{1}{a_i} = \frac{1}{a_{k-2}} + \frac{1}{a_{k-1}} + \frac{1}{a_k}. \tag{2.20}$$

We suppose the denominators a_i in equation (2.19) are given in increasing order and prove upper bounds for the size of a_i, $1 \leq i \leq k$. In particular we use an induction argument to show that $a_i \leq \alpha_i n^{2^{i-1}}$ where the finite sequence $(\alpha_i)_{1 \leq i \leq k}$ is recursively defined by $\alpha_1 = k$ and $\alpha_i = (k - i + 1) \prod_{j<i} \alpha_j$ for $2 \leq i \leq k$. For $i = 1$ this bound follows easily from the following inequality

$$\frac{m}{n} = \frac{1}{a_1} + \cdots + \frac{1}{a_k} \leq \frac{k}{a_1}$$

which leads to $a_1 \leq \frac{kn}{m} \leq kn$. If we suppose the bound holds for a_i, with a similar argument we get

$$\frac{m}{n} - \frac{1}{a_1} - \cdots - \frac{1}{a_i} = \frac{1}{a_{i+1}} + \cdots + \frac{1}{a_k} \leq \frac{(k-i)}{a_{i+1}}.$$

The last inequality together with the induction hypothesis for $j < i + 1$ implies

$$a_{i+1} \leq (k-i) \frac{n \prod_{j<i+1} a_j}{m \prod_{j<i+1} a_j - n \sum_{j<i+1} \prod_{\substack{l<i+1 \\ l \neq j}} a_l} \leq (k-i)n \prod_{j<i+1} a_j \leq \alpha_{i+1} n^{2^i}.$$

By definition α_i is a polynomial in k of degree 2^i with leading coefficient 1. Furthermore, the denominator of the rational number on the left hand side of equation (2.20) is of size at most $n \prod_{i=1}^{k-3} a_i \ll_k n^{2^{k-3}}$. By the aforementioned result we can compute all decompositions as a sum of three unit fractions of this number in time $\mathcal{O}_{\epsilon,k}(n^{2^{k-3}(3/5+\epsilon)})$. We have to compute these representations for at most $\prod_{i=1}^{k-3} a_i \ll_k n^{2^{k-3}-1}$ rational numbers which leads to an upper bound of

$$\mathcal{O}_{\epsilon,k}\left(n^{2^{k-3}(8/5+\epsilon)-1}\right)$$

for the running time. $\square$

Remark 2.13. *The procedure for computing representations as a sum of k unit fractions as described in the proof of Corollary 2.4 could lead to a speedup for calculations similar to those in* [2]*. In the calculations above the size of the numerator of the rational number on the left hand side of equation* (2.20)*, which we denote by* $\frac{m'}{n'}$*, was not taken into* *account. We note that also the proof of the upper bound for $f_3(m,n)$ by Browning and Elsholtz* [9, Theorem 2] *may be similarly turned into an algorithm of expected running time $\mathcal{O}_\epsilon\big(n^\epsilon\big(\frac{n}{m}\big)^{2/3}\big)$. In practice one would check dynamically if $m' \ll (n')^{1/4}$ before computing the representations as a sum of three unit fractions of $\frac{m'}{n'}$. If this is the case, the algorithm described in the first part of the proof of Corollary 2.4 should be applied, if $m' \gg (n')^{1/4}$ the method of* [9] *should be used.*

2.6. Sums of k unit fractions

In this section we will prove Theorem 2.5. Browning and Elsholtz used an induction argument on their bound for the quantity $f_3(m,n)$ to get bounds for $f_k(m,n)$ for $k \geq 4$. Using their arguments directly on our result from Theorem 2.2 would lead to worse upper bounds than those of Browning and Elsholtz. The reason is that our bound for $f_3(m,n)$ is weaker than the one in [9] when m is large.

As in [9, Section 4] the proof of Theorem 2.5 will be based on the observation that from equation (2.5) it follows that

$$f_k(m,n) \leq \sum_{\frac{n}{m} < n_1 t_1 \leq \frac{kn}{m}} f_{k-1}(mn_1t_1 - n, n_1t_1n),$$

which, after introducing the parameter $u = mn_1t_1 - n$, becomes

$$f_k(m,n) \leq \sum_{\substack{0<u\leq(k-1)n \\ m \mid u+n}} f_{k-1}\left(u, \frac{n(u+n)}{m}\right). \tag{2.21}$$

The improvement in Theorem 2.5 stems from extending the method of Browning and Elsholtz by applying the following new idea. In the case of $k = 4$ we do not consider the sum on the right hand side of (2.21) as a whole but we split the sum into two parts. In the first part we collect the values of u where $0 < u \leq n^\delta$ for some $0 < \delta < 1$ which will be chosen later. This sum will be small since it contains few summands.

The second part will consist of all summands where $u > n^\delta$. This corresponds to $n_1t_1 > \frac{n+n^\delta}{m}$ which will force n_2t_2 and n_3t_3 to be small.

The following Lemma 2.14 is [9, Theorem 2].

Lemma 2.14 (Browning, Elsholtz (2011))**.** *For any $\epsilon > 0$, we have*

$$f_3(m,n) \ll_\epsilon n^\epsilon \left(\frac{n}{m}\right)^{2/3}.$$

In the proof of Theorem 2.5 below we make use of Lemma 2.14 rather than Theorem 2.2. Furthermore, we will use a lifting procedure which was first used by Browning and Elsholtz [9, Section 4] to lift upper bounds of the form

$$f_5(m,n) \ll_\epsilon n^\epsilon \left(\frac{n^2}{m}\right)^c \tag{2.22}$$

to upper bounds for $f_k(m,n)$ for $k > 5$. For possible future use we write this procedure up in the following lemma and work through the original proof by Browning and Elsholtz with an arbitrary exponent $c > 1$ in (2.22).

Lemma 2.15 (Browning, Elsholtz (2011))**.** *Suppose that there exists $c > 1$ such that*

$$f_5(m,n) \ll_\epsilon n^\epsilon \left(\frac{n^2}{m}\right)^c.$$

Then for any $k \geq 5$ we have

$$f_k(m,n) \ll_\epsilon (kn)^\epsilon \left(\frac{k^{4/3}n^2}{m}\right)^{c2^{k-5}}.$$

Proof. We will inductively show that for $k \geq 5$ there exists Θ_k depending on k such that we have

$$f_k(m,n) \ll_\epsilon (kn)^\epsilon \left(\frac{k^{\Theta_k}n^2}{m}\right)^{c2^{k-5}} \tag{2.23}$$

and we note that this is certainly true for $k = 5$ by assumption. The proof works in three steps.

1. Establish an upper bound where the implied constant is allowed to depend on k. For $k \geq 5$ we want to have a bound of the form

$$f_k(m,n) \ll_{k,\epsilon} n^\epsilon \left(\frac{n^2}{m}\right)^{c2^{k-5}} \tag{2.24}$$

where the implied constant is allowed to depend on k. An upper bound of this type may easily be achieved via (2.21). Indeed this bound holds true for $k = 5$ by assumption and assuming its existence for $f_k(m,n)$ we find for $f_{k+1}(m,n)$

$$f_{k+1}(m,n) \ll \sum_{\substack{0<u\leq kn \\ m|u+n}} f_k\left(u, \frac{n(u+n)}{m}\right) \ll_{k,\epsilon} n^\epsilon \left(\frac{n^2}{m}\right)^{c2^{k-4}} \sum_{u=1}^{\infty} \frac{1}{u^{c2^{k-5}}}$$

$$\ll_{k,\epsilon} n^{\epsilon} \left(\frac{n^2}{m}\right)^{c2^{k-4}},$$

where we used that $c > 1$.

2. Use inequality (2.21) and split the sum into two parts.

For the upper bound where the implied constant is independent of k we again suppose it to be true for $f_k(m,n)$ with $k \geq 5$ and inductively prove it to hold for $f_{k+1}(m,n)$. Using inequalities (2.21) and (2.23) we get

$$\begin{aligned}
f_{k+1}(m,n) &\ll \sum_{\substack{0<u\leq kn \\ m|u+n}} f_k\left(u, \frac{n(u+n)}{m}\right) \\
&\ll \sum_{\substack{0<u\leq (L-1)n \\ m|u+n}} f_k\left(u, \frac{n(u+n)}{m}\right) + \sum_{\substack{(L-1)n<u\leq kn \\ m|u+n}} f_k\left(u, \frac{n(u+n)}{m}\right) \\
&\ll (kn)^{\epsilon} k^{\Theta_k c2^{k-5}} \left(\frac{n^2}{m}\right)^{c2^{k-4}} \times \\
&\left(\sum_{0<u\leq (L-1)n} \frac{1}{u^{c2^{k-5}}} L^{c2^{k-4}} + \sum_{(L-1)n<u\leq kn} \frac{1}{u^{c2^{k-5}}} (k+1)^{c2^{k-4}}\right).
\end{aligned}$$

Since $c2^{k-5} > 1$ the infinite sums over $\frac{1}{u^{c2^{k-5}}}$ converge. For the first sum we use that the sum is bounded by a constant for the second sum we use the following more accurate bound

$$\sum_{(L-1)n<u\leq kn} \frac{1}{u^{c2^{k-5}}} \leq \sum_{u=L}^{\infty} \frac{1}{u^{c2^{k-5}}} \ll \int_L^{\infty} \frac{1}{u^{c2^{k-5}}} \mathrm{d}u \ll L^{1-c2^{k-5}}.$$

Together with the fact that $(a+b)^{\alpha} \geq a^{\alpha} + b^{\alpha}$ for $a, b > 0$ and $\alpha > 1$ this shows that

$$\begin{aligned}
&f_{k+1}(m,n) \\
&\quad \ll_{\epsilon} ((k+1)n)^{\epsilon} (k+1)^{\Theta_k c2^{k-5}} \left(\frac{n^2}{m}\right)^{c2^{k-4}} \left(L^{c2^{k-4}} + \left(\frac{k+1}{L^{1/2-(c2^{k-4})^{-1}}}\right)^{c2^{k-4}}\right) \\
&\quad \ll_{\epsilon} ((k+1)n)^{\epsilon} (k+1)^{\Theta_k c2^{k-5}} \left(\frac{n^2}{m}\right)^{c2^{k-4}} \left(L + \frac{k+1}{L^{1/2-(c2^{k-4})^{-1}}}\right)^{c2^{k-4}}.
\end{aligned}$$

3. Optimizing for L and determining an upper bound for Θ_k.

By the bound we derived in step 1 we may suppose that $k \geq \max\{\frac{\log(\frac{2}{3}(c\epsilon)^{-1})}{\log 2} +$

$4, (\frac{1+\sqrt{5}}{2})^{1/\epsilon} - 1\}$. With $L = (k+1)^{2/3}$ we get

$$\begin{aligned} & f_{k+1}(m,n) \\ & \ll_\epsilon ((k+1)n)^\epsilon (k+1)^{\Theta_k c 2^{k-5}} \left(\frac{n^2}{m}\right)^{c2^{k-4}} (k+1)^{2/3 \cdot c2^{k-4}} \left(1 + L^{(c2^{k-4})^{-1}}\right)^{c2^{k-4}} \\ & \ll_\epsilon (k+1)^{\epsilon(1+c2^{k-3})} n^\epsilon (k+1)^{c2^{k-4}(\Theta_k/2+2/3)} \left(\frac{n^2}{m}\right)^{c2^{k-4}}. \end{aligned}$$

With $\Theta_{k+1} = \frac{\Theta_k}{2} + \frac{2}{3}$ and an appropriate choice of ϵ this implies

$$f_{k+1} \ll_\epsilon ((k+1)n)^\epsilon \left(\frac{(k+1)^{\Theta_{k+1}} n^2}{m}\right)^{c2^{(k+1)-5}}$$

Since for $\Theta_5 \leq \frac{4}{3}$ the sequence recursively defined by $\Theta_{k+1} = \frac{\Theta_k}{2} + \frac{2}{3}$ monotonically increases towards its limit $\frac{4}{3}$ we eventually get for any $k \geq 5$:

$$f_k(m,n) \ll_\epsilon (kn)^\epsilon \left(\frac{k^{4/3} n^2}{m}\right)^{c2^{k-5}}.$$

□

Proof of Theorem 2.5. In the following $\delta < 1$ is a fixed constant to be chosen at the end of the proof. We start with proving bounds on $f_4(m,n)$ and we write $f_4(m,n) = f_4^{(1)}(m,n) + f_4^{(2)}(m,n)$. Here $f_4^{(1)}(m,n)$ counts those solutions of equation (2.5) with $n_1 t_1 \leq \frac{n+n^\delta}{m}$ and $f_4^{(2)}(m,n)$ those with $n_1 t_1 > \frac{n+n^\delta}{m}$. From (2.21) we have that

$$\begin{aligned} f_4(m,n) = f_4^{(1)}(m,n) + f_4^{(2)}(m,n) &\leq \sum_{\substack{0<u\leq n^\delta \\ m|u+n}} f_3\left(u, \frac{n(u+n)}{m}\right) + f_4^{(2)}(m,n) \\ &= S_1 + f_4^{(2)}(m,n). \end{aligned}$$

We use the following estimate (uniform in $a \in \mathbb{Z}$)

$$\sum_{\substack{n\leq x \\ n\equiv a \bmod q}} n^{-\Theta} = \frac{x^{1-\Theta}}{(1-\Theta)q} + \mathcal{O}_\Theta(1). \tag{2.25}$$

To bound the sum S_1 we use (2.25) and Lemma 2.14 to get

$$S_1 \ll_\epsilon n^\epsilon \left(\frac{n^2}{m}\right)^{\frac{2}{3}} \sum_{\substack{0<u\leq n^\delta \\ m|u+n}} \frac{1}{u^{\frac{2}{3}}} \ll_\epsilon n^\epsilon \left(\frac{n^2}{m}\right)^{\frac{2}{3}} \left(\frac{n^{\frac{\delta}{3}}}{m} + 1\right). \tag{2.26}$$

Next we prove that

$$f_4^{(2)}(m,n) \ll_\epsilon n^\epsilon \frac{n^{(12-4\delta)/5}}{m^{8/5}}.$$

Since there are at most $\mathcal{O}_\epsilon(n^\epsilon)$ distinct patterns (n_1,n_2,n_3,n_4) it suffices to prove this bound for all solutions counted by $f_4^{(2)}(m,n)$ corresponding to a fixed pattern. To get an upper bound for the contribution of $f_4^{(2)}(m,n)$ we thus suppose that (n_1,n_2,n_3,n_4) is fixed and note that the fact that $\frac{4n}{m} \geq n_1t_1 > \frac{n+n^\delta}{m}$ implies the following upper bound for n_2t_2:

$$\frac{3}{n_2t_2} \geq \frac{mn_1t_1 - n}{nn_1t_1} \geq \frac{mn^\delta}{4n^2}.$$

Therefore, we have

$$n_2t_2 \ll \frac{n^{2-\delta}}{m}. \tag{2.27}$$

We use again relative greatest common divisors and write a representation of $\frac{m}{n}$ as a sum of four unit fractions as

$$\begin{aligned}\frac{m}{n} = &\frac{1}{n_1x_1x_{12}x_{13}x_{14}x_{123}x_{124}x_{134}x_{1234}} + \frac{1}{n_2x_2x_{12}x_{23}x_{24}x_{123}x_{124}x_{234}x_{1234}} \\ &+ \frac{1}{n_3x_3x_{13}x_{23}x_{34}x_{123}x_{134}x_{234}x_{1234}} + \frac{1}{n_4x_4x_{14}x_{24}x_{34}x_{124}x_{134}x_{234}x_{1234}}.\end{aligned}$$

It is again easy to see that $x_1 = x_2 = x_3 = x_4 = 1$ and multiplying out the last equation yields

$$\begin{aligned}&mx_{12}x_{13}x_{14}x_{23}x_{24}x_{34}x_{123}x_{124}x_{134}x_{234}x_{1234} \\ &= \frac{n}{n_1}x_{23}x_{24}x_{34}x_{234} + \frac{n}{n_2}x_{13}x_{14}x_{34}x_{134} + \frac{n}{n_3}x_{12}x_{14}x_{24}x_{124} + \frac{n}{n_4}x_{12}x_{13}x_{23}x_{123}.\end{aligned} \tag{2.28}$$

From equation (2.28) we see that the quantity

$$z_{34} = \frac{\frac{n}{n_3}x_{12}x_{14}x_{24}x_{124} + \frac{n}{n_4}x_{12}x_{13}x_{23}x_{123}}{x_{34}}$$

is an integer and we use

$$z_{34}x_{34} = \frac{n}{n_3}x_{12}x_{14}x_{24}x_{124} + \frac{n}{n_4}x_{12}x_{13}x_{23}x_{123}. \tag{2.29}$$

By (2.27) and $\frac{4n}{m} \geq n_1t_1 > \frac{n+n^\delta}{m}$ we have

$$(t_1t_2)^4 = (x_{12}x_{13}x_{14}x_{123}x_{124}x_{134}x_{1234})^4(x_{12}x_{23}x_{24}x_{123}x_{124}x_{234}x_{1234})^4 \ll \frac{n^{12-4\delta}}{m^8}, \tag{2.30}$$

and we write

$$\begin{aligned}
&(x_{12}x_{13}x_{14}x_{123}x_{124}x_{134}x_{1234})^4(x_{12}x_{23}x_{24}x_{123}x_{124}x_{234}x_{1234})^4 = \\
&(x_{12}x_{13}x_{14}x_{23}x_{24}x_{123}x_{124}x_{1234})(x_{12}x_{13}x_{23}x_{24}x_{123}x_{124}x_{134}x_{234}x_{1234})\times \\
&(x_{12}x_{14}x_{23}x_{24}x_{123}x_{124}x_{134}x_{234}x_{1234})(x_{12}x_{13}x_{14}x_{24}x_{123}x_{124}x_{134}x_{234}x_{1234})\times \\
&(x_{12}^4x_{13}x_{14}x_{23}x_{123}^4x_{124}^4x_{134}x_{234}x_{1234}^4).
\end{aligned} \tag{2.31}$$

We show that each of the five factors in brackets on the right hand side of the last equation corresponds to at most $\mathcal{O}_\epsilon(n^\epsilon)$ solutions of (2.28), where ϵ is an arbitrarily small positive number. First we note that all factors are of polynomial size in n and by Lemma 2.12, given one of these factors, we have $\mathcal{O}_\epsilon(n^\epsilon)$ choices for all the x_{ij}, x_{ijk} and x_{1234} appearing as sub-factors.

Given positive integer constants C_0, C_1, C_2 and C_3 of size polynomial in n, we count the number of integer solutions (A, B) of the equation

$$C_0AB = C_1A + C_2B + C_3. \tag{2.32}$$

Rewriting this equation in the form

$$(C_0A - C_2)(C_0B - C_1) = C_0C_3 + C_1C_2$$

we see that the number of solutions (A, B) is bounded by $\mathcal{O}_\epsilon(n^\epsilon)$. For the second to the fifth factor on the right hand side of (2.31) exactly two parameters are missing to uniquely determine a solution of (2.28). All of these factors miss the parameter x_{34}. The second one additionally misses x_{14}, the third one x_{13}, the fourth one x_{23} and the last one x_{24}. In all of these cases equation (2.28) provides an instance of (2.32) where the variables A and B correspond to the two missing parameters (the term containing both missing parameters on the right hand side of (2.28) may be shifted to the left hand side).

In the first factor on the right hand side of (2.31) three parameters are missing. From equation (2.29) we see that we have at most $\mathcal{O}_\epsilon(n^\epsilon)$ choices for the parameter x_{34}. To see the same bound for the parameters x_{134} and x_{234} we use again that equations of type (2.32) can be factorized.

Since by (2.30) at least one of the factors on the right hand side of (2.31) is of order $\mathcal{O}\big(\frac{n^{(12-4\delta)/5}}{m^{8/5}}\big)$ we have that

$$f_4^{(2)}(m,n) \ll_\epsilon n^\epsilon \frac{n^{(12-4\delta)/5}}{m^{8/5}}. \tag{2.33}$$

Again we note that in the considerations above the divisor bound from Lemma 2.12 was applied a bounded number of times and the bound in (2.33) follows upon redefining the choice of ϵ. Choosing $\delta = \frac{16}{17}$ in (2.26) and (2.33) we get

$$f_4(m,n) \ll n^\epsilon \left(\frac{n^{4/3}}{m^{2/3}} + \frac{n^{28/17}}{m^{8/5}} \right). \tag{2.34}$$

To bound $f_5(m,n)$ we again use (2.21) and (2.25) and get

$$f_5(m,n) \ll n^\epsilon \sum_{\substack{0<u\le 4n \\ m|u+n}} \left(\left(\frac{n^2}{m}\right)^{4/3} \frac{1}{u^{2/3}} + \left(\frac{n^2}{m}\right)^{28/17} \frac{1}{u^{8/5}} \right) \ll n^\epsilon \left(\frac{n^2}{m}\right)^{28/17}. \tag{2.35}$$

Setting $c = \frac{28}{17}$ in Lemma 2.15 yields the bound in Theorem 2.5. □

2.7. Lower bounds

Proof of Theorem 2.7. To prove the first bound we are going to extend an idea used in the proof of [9, Theorem 1]. As before we use highly composite denominators $n \in \mathbb{N}$, but here we show that there are many values a_1 with many corresponding pairs (a_2, a_3) giving a solution of

$$\frac{m}{n} = \frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3}.$$

To prove our lower bound for $f_3(m,n)$ we consider the set

$$\mathcal{N} = \left\{ mn' : n' = \prod_{i=1}^{r} p_i \right\},$$

where p_i is the i-th prime. In choosing the denominators $n \in \mathcal{N}$ we reduce the problem to finding many solutions of the equation

$$\frac{1}{n'} = \frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3}.$$

We set $a_1 = n' + d$, where d is any divisor of n', and are left with

$$\frac{1}{n'} - \frac{1}{n'+d} = \frac{1}{n'\left(\frac{n'}{d}+1\right)} = \frac{1}{a_2} + \frac{1}{a_3}.$$

For two divisors d_1 and d_2 of n' with $(d_1, d_2) = 1$ we have

$$\frac{1}{n'\left(\frac{n'}{d}+1\right)} = \frac{1}{\frac{n'\left(n'/d+1\right)}{d_1}(d_1+d_2)} + \frac{1}{\frac{n'\left(n'/d+1\right)}{d_2}(d_1+d_2)}. \tag{2.36}$$

We note that for two pairs of divisors d_1, d_2 and d_1', d_2' with $(d_1, d_2) = 1$ and $(d_1', d_2') = 1$ it follows that

$$\frac{n'\left(\frac{n'}{d}+1\right)}{d_1}(d_1+d_2) = \frac{n'\left(\frac{n'}{d}+1\right)}{d_1'}(d_1'+d_2') \Leftrightarrow \frac{d_1}{d_2} = \frac{d_1'}{d_2'}.$$

Since d_1 and d_2 as well as d_1' and d_2' are coprime we get $d_1 = d_1'$ and $d_2 = d_2'$. This implies that each pair (d_1, d_2) with $d_1 < d_2$ gives a unique solution of equation (2.36). Furthermore, for any choice of d, d_1, d_2 it follows that

$$n' + d < \frac{n'\left(\frac{n'}{d}+1\right)}{d_2}(d_1+d_2),$$

which altogether implies that by counting all possible choices for d, d_1, d_2 we get a lower bound for twice the value of $f_3(1, n')$.

Choosing n' as in the construction of the set $\mathcal{N}$, we have $2^{\omega(n')}$ choices for the divisor d and using the binomial theorem there are

$$\sum_{i=0}^{\omega(n')} \binom{\omega(n')}{i} \sum_{j=0}^{\omega(n')-i} \binom{\omega(n')-i}{j} = \sum_{i=0}^{\omega(n')} \binom{\omega(n')}{i} 2^{\omega(n')-i} = 3^{\omega(n')}$$

choices for the divisors d_1 and d_2. As a consequence of the prime number theorem it is known that $\omega(n') \sim \frac{\log n'}{\log\log n'}$ and hence, for $n \in \mathcal{N}$

$$f_3(m,n) = f_3(1,n') \geq \frac{1}{2} 2^{\omega(n')} 3^{\omega(n')} \geq \exp\left((\log 6 + o(1)) \frac{\log n'}{\log\log n'}\right)$$

$$\geq \exp\left((\log 6 + o_m(1)) \frac{\log n}{\log\log n} \right).$$

For the second bound we modify the idea used in the proof of [25, Theorem 1.8]. For fixed $m \in \mathbb{N}$, as a consequence of the Turán-Kubilius inequality (see e.g. [73, p. 434]) we get that the set

$$\mathcal{M}_1 = \bigcap_{\substack{k \leq m \\ (k,m)=1}} \left\{ n \in \mathbb{N} : \omega(n,k,m) = \left(\frac{1}{\varphi(m)} + o(1) \right) \log\log n \right\}$$

is a set with density one, i.e. $\lim_{x\to\infty} \frac{\{n \in \mathcal{M}_1 : n \leq x\}}{x} = 1$.

For any $n \in \mathcal{M}_1$ we write $\frac{m}{n} = \frac{m'}{n'}$ with $(m', n') = 1$ and note that $\omega(n,k,m) = \omega(n',k,m)$ for all k with $(k,m) = 1$. By construction of the set $\mathcal{M}_1$ and since n' is coprime to m', we find $\left(\frac{1}{\varphi(m)} + o(1)\right) \log\log n$ prime divisors p of n' in the residue class $-n' \bmod m'$. For any of these prime divisors we have

$$\frac{m'}{n'} - \frac{1}{\frac{n'+p}{m'}} = \frac{p}{n' \frac{n'+p}{m'}} = \frac{1}{n' \frac{n'/p+1}{m'}}$$

where $\frac{n'/p+1}{m'}$ is an integer. Again, by construction of the set $\mathcal{M}_1$, for the number of prime factors of n' we have

$$\omega(n') \geq \omega(n) - \omega(m) = (1 + o_m(1)) \log\log n.$$

For two coprime divisors d_1 and d_2 of n' we construct decompositions of $\frac{1}{n' \frac{n'/p+1}{m'}}$ as a sum of two unit fractions as in (2.36). As above we see that for any prime divisor p of n' in the residue class $-n' \bmod m'$ there are at least $3^{\omega(n')}$ such decompositions and all of them are distinct.

Altogether this implies that for any $n \in \mathcal{M}_1$

$$\begin{aligned} f(m,n) &\geq \left(\frac{1}{\varphi(m)} + o(1) \right) 3^{\omega(n')} \cdot \log\log n \geq \left(\frac{1}{\varphi(m)} + o(1) \right) 3^{\omega(n/m)} \cdot \log\log n \\ &\geq \exp((\log 3 + o_m(1)) \log\log n) \cdot \log\log n. \end{aligned}$$

Finally, we prove the improved lower bound on $f_3(4,n)$. To do so, we set

$$\mathcal{M}_2 = \left(\bigcap_{i\in\{1,3\}} \{ n \in \mathbb{N} : \frac{\tau(n,4)}{4} \leq \tau(n,i,4) \} \right) \cap$$

$$\cap \{n \in \mathbb{N} : \omega(n) = (1+o(1)) \log\log n\} \cap \{n \in \mathbb{N} : \tau(n) \geq (\log n)^{\log 2 + o(1)}\}.$$

The first two sets with $i = 1$ and $i = 3$ in the intersection in the definition of $\mathcal{M}_2$ have density 1 by [46, Theorem 5]. For the third and the fourth set this is true by the Turán-Kubilius inequality (again see e.g. [73, p. 434]). Hence the set $\mathcal{M}_2$ has density 1 and we investigate what happens for n in a certain residue class modulo 4.

If $n \equiv 0 \bmod 4$, then $\frac{4}{n} = \frac{1}{n/4}$ and for any divisor d of $\frac{n}{4}$ we have

$$\frac{1}{\frac{n}{4}} - \frac{1}{\frac{n}{4} + d} = \frac{1}{\frac{n}{4}\left(\frac{n}{4d} + 1\right)}.$$

Since $\omega\left(\frac{n}{4}\right) \geq \omega(n) - 1$, with the same arguments as above, we conclude that the number of representations of $\frac{1}{n/4(n/(4d)+1)}$ as a sum of two unit fractions is at least of order $3^{\omega(n/4)} = 3^{(1+o(1))\log\log n}$. From $\tau(n) = \prod_{p|n}(\nu_p(n)+1)$ we easily deduce that $\tau\left(\frac{n}{4}\right) \geq \frac{1}{3}\tau(n)$. Altogether we thus get

$$f_3(4,n) \geq \frac{1}{3}\tau\left(\frac{n}{4}\right) 3^{\omega(n/4)} \geq \exp((\log 6 + o(1)) \log\log n).$$

If $n \equiv 2 \bmod 4$, then $\frac{n}{2}$ is odd and the same is true for all $\tau\left(\frac{n}{2}\right) = \frac{1}{2}\tau(n)$ divisors of $\frac{n}{2}$. We have $\frac{4}{n} = \frac{2}{n/2}$ and for any divisor d of $\frac{n}{2}$

$$\frac{2}{\frac{n}{2}} - \frac{1}{\frac{n/2+d}{2}} = \frac{1}{\frac{n}{2}\left(\frac{n/2d+1}{2}\right)}.$$

As above we get

$$f_3(4,n) \geq \tau\left(\frac{n}{2}\right) 3^{\omega(n)-1} \geq \exp((\log 6 + o(1)) \log\log n).$$

Finally, if $n \equiv r \bmod 4$ for $r \in \{1,3\}$, we have $\tau(n,4) = \tau(n)$ and by construction of the set $\mathcal{M}_2$, we have more than $\frac{\tau(n)}{4}$ divisors d of n in the residue class $-r \bmod 4$. Again, for any of these divisors we have

$$\frac{4}{n} - \frac{1}{\frac{n+d}{4}} = \frac{1}{n\left(\frac{n/d+1}{4}\right)}.$$

Applying the arguments used previously one more time, we find

$$f_3(4,n) \geq \frac{\tau(n)}{4} 3^{\omega(n)} \geq \exp((\log 6 + o(1)) \log\log n)$$

also in this case.

$\square$

Remark 2.16. *The difference in the constants in the exponential functions of the lower bounds on $f(m,n)$ and $f(4,n)$ for sets of integers with density one in Theorem 2.7 is basically due to cancellation effects when dealing with general m. In particular we deal with $\frac{m}{n} = \frac{m'}{n'}$, where $(m',n') = 1$, and we would need to have good control of the number of divisors of n' in the residue class $-n' \bmod m'$ to get the $\log 6$ exponent also in the general case. However, if we do not ask about a lower bound holding for a set of density one within the positive integers, but for a set of integers of density one within the set $\mathcal{S}$ of positive integers coprime to a given $m \in \mathbb{N}$, we may achieve the $\log 6$ exponent. To do so we replace the set $\mathcal{M}_1$ with*

$$\mathcal{M}_1' = \left(\bigcap_{\substack{1 \le i \le m \\ (i,m)=1}} \{n \in \mathbb{N} : \tau(n,i,m) = \frac{\tau(n)}{\varphi(m)}(1 + o_m(1))\} \right) \cap$$
$$\cap \{n \in \mathbb{N} : \omega(n) = (1 + o(1)) \log\log n\} \cap \{n \in \mathbb{N} : \tau(n) \ge (\log n)^{\log 2 + o(1)}\} \cap \mathcal{S}.$$

Now we may use results from [46, Theorem 5] *as well as Turán-Kubilius like previously and get that $\mathcal{M}_1'$ has density one in $\mathcal{S}$. Instead of constructing the first denominator via shifts in prime factors of n we may use arbitrary divisors of n in this case, which leads to the improvement mentioned above.*

Proof of Theorem 2.10. We consider solutions corresponding to the pattern $(1,p,p)$. In equation (2.1) we suppose that a_1 is the denominator with $(a_1,p) = 1$ and we write $a_1 = t_1$, $a_2 = pt_2$ and $a_3 = pt_3$. We use the parametrization via relative greatest common divisors of the t_i and applying Lemma 2.11 it is easy to see, that $x_1 = x_2 = x_3 = 1$ in this case. Hence we are looking for infinitely many primes $p \equiv e \bmod f$ such that for given $m \in \mathbb{N}$ the equation

$$\frac{m}{p} = \frac{1}{x_{12}x_{13}x_{123}} + \frac{1}{px_{12}x_{23}x_{123}} + \frac{1}{px_{13}x_{23}x_{123}} \tag{2.37}$$

has many solutions. Multiplying equation (2.37) by the common denominator we get

$$mx_{12}x_{13}x_{23}x_{123} = px_{23} + x_{13} + x_{12}.$$

Setting $x_{12}+x_{13}=kx_{23}$, $M=\operatorname{lcm}(m,f)$ and $x_{12}=\frac{M}{m}$ we deduce that

$$M\left(kx_{23}-\frac{M}{m}\right)x_{123}=p+k.$$

The residue class $(f-e)\equiv -e \bmod f$ splits into the residue classes $(f-e)+if \bmod M$, for $0\le i\le \frac{m}{(m,f)}-1$. Note, that $\gcd\left(f,\frac{m}{(m,f)}\right)=1$ hence the integers $i\cdot f$ for $0\le i\le \frac{m}{(m,f)}-1$ are a full system of residues modulo $\frac{m}{(m,f)}$. In particular there exists a $0\le j\le \frac{m}{(m,f)}-1$ such that $(f-e)+jf\equiv 1 \bmod \frac{m}{(m,f)}$. We set $k=(f-e)+jf$ and with $(e,f)=1$ we altogether see that $(M,k)=1$.

Now let $Q=\prod_{i=1}^{r}q_i$ where q_i is the i-th prime with $q_i\equiv -\frac{M}{m} \bmod k$ and $q_i>M$. Note that $\gcd(M,Q)=1$.

With $r=\left\lfloor\frac{\log t}{\varphi(k)C\log\log t}\right\rfloor$ we find that Q is of order $t^{1/C+o_{f,m}(1)}$. We now use Linnik's theorem on primes in arithmetic progressions. As the modulus is very smooth[3] we can use an exponent of $C=\frac{12}{5}+o(1)$, due to Chang [11, Corollary 11]. Hence we may find a prime p of order $M^C t^{1+o_{f,m}(1)}$ with

$$p\equiv -k \bmod QM.$$

This congruence implies that $p+k$ is divisible by the primes $q_1,\dots,q_r$ and together with $k=(f-e)+jf$, we deduce that $p\equiv e \bmod f$ and $p+k\equiv 0 \bmod M$.

Let $l\in\mathbb{N}_0$ and S be a subset of size $l\operatorname{ord}_k\left(-\frac{M}{m}\right)+1$ of the prime factors of Q. Hence $x_{23}=\frac{\prod_{q\in S}q+\frac{M}{m}}{k}$ is an integer and we set $x_{123}=\frac{p+k}{M\prod_{q\in S}q}$. We observe that any of these choices leads to a different solution of (2.37). To see this we look at the denominator $a_2=px_{12}x_{23}x_{123}$ of the second fraction on the right hand side of this equation. Suppose that two sets S and S' would lead to the same denominator a_2. With $x_{12}=\frac{M}{m}$ this would imply the existence of $x_{23}\neq x'_{23}$ such that

$$p\frac{M}{m}x_{23}\frac{p+k}{M(kx_{23}-\frac{M}{m})}=p\frac{M}{m}x'_{23}\frac{p+k}{M(kx'_{23}-\frac{M}{m})}$$

from which we derive that

$$\frac{x_{23}}{x'_{23}}=\frac{kx_{23}-\frac{M}{m}}{kx'_{23}-\frac{M}{m}}=\frac{\prod_{q\in S}q}{\prod_{q'\in S'}q'}.$$

If $q\in S$ would divide x_{23} then q would also divide $\frac{M}{m}$, which is impossible by construction

[3]We note that more recently instead of the term *smooth numbers* also the term *friable numbers* is used for integers without large prime factors.

of Q. We hence have that $\frac{\prod_{q\in S} q}{\prod_{q'\in S'} q'} = 1$ and thus $S = S'$.

To count the number of solutions we get with the above construction, we make use of a formula which can be found in [6, Theorem 1], for example, and which states

$$\sum_{i\geq 0} \binom{n}{iu} = \frac{1}{u}\sum_{j=0}^{u-1}(1+\xi_u^j)^n, \tag{2.38}$$

where $\xi_u = \exp\left(\frac{2\pi i}{u}\right)$. Note that for the term corresponding to $j = 0$ in the sum on the right hand side of (2.38) we get 2^n while for all other j we have $|1+\xi_u^j| < 2$. Hence we deduce

$$\sum_{i\geq 0} \binom{n}{iu} = \frac{2^n}{u}(1+o_u(1)).$$

The number of choices of the parameter x_{23} is

$$\begin{aligned}\sum_{i\geq 0} \binom{r}{i\operatorname{ord}_k\left(-\frac{M}{m}\right)+1} &= \sum_{i\geq 0} \frac{r}{i\operatorname{ord}_k\left(-\frac{M}{m}\right)+1}\binom{r-1}{i\operatorname{ord}_k\left(-\frac{M}{m}\right)} \geq \sum_{i\geq 0}\binom{r-1}{i\operatorname{ord}_k\left(-\frac{M}{m}\right)} \\ &= \frac{2^{r-1}}{\operatorname{ord}_k\left(-\frac{M}{m}\right)}(1+o_{f,m}(1)).\end{aligned}$$

Plugging in $r = \left\lfloor \frac{\log t}{\varphi(k)C\log\log t}\right\rfloor$ and using that $p \leq M^C t^{1+o_{f,m}(1)}$ we get a lower bound of

$$\begin{aligned}f_3(m,p) &\gg_{f,m} \exp\left(\left(\frac{\log 2}{C\varphi(k)} + o_{f,m}(1)\right)\frac{\log t}{\log\log t}\right) \\ &\gg_{f,m} \exp\left(\left(\frac{5\log 2}{12\operatorname{lcm}(m,f)} + o_{f,m}(1)\right)\frac{\log p}{\log\log p}\right).\end{aligned} \tag{2.39}$$

□

Remark 2.17. *The best known exponent for Linnik's Theorem takes care of the worst case modulus and is* 5 *by work of Xylouris* [75]*. Chang's result* [11, Corollary 11] *considers smooth moduli (as in our situation) and allows for the better exponent* $\frac{12}{5}+o(1)$*. Harman investigated, in connection with constructing Carmichael numbers, what happens if one is allowed to avoid a small set of exceptional moduli. In this situation he improved the exponent to* $\frac{1}{0.4736}$ *(see* [49, Theorem 1.2] *and* [48] *for some more explanation). As in our situation we choose the modulus* MQ*, and hence can avoid "bad" factors, it seems possible that Theorem 2.10 can also be proved with a factor of* 0.4736 *instead of* $\frac{5}{12} = 0.4166\ldots$ *in the exponent of the lower bound on* $f_3(m,p)$*.*

Remark 2.18. *If we consider the case* $m = 4$*,* $f = 4$ *and* $e \in \{1,3\}$ *in Theorem 2.10,* *we can explicitly compute k in the first line of* (2.39). *We simply have $k = 3$ if $e = 1$ and $k = 1$ if $e = 3$ hence we arrive at the lower bounds*

$$f_3(4,p) \gg \exp\left((0.1444 + o(1))\frac{\log p}{\log\log p}\right)$$

if $e = 1$ and

$$f_3(4,p) \gg \exp\left((0.2888 + o(1))\frac{\log p}{\log\log p}\right)$$

if $e = 3$.

2.8. Acknowledgements

The authors acknowledge the support of the Austrian Science Fund (FWF): W1230. Furthermore, we would like to thank Igor Shparlinski and Glyn Harman for drawing our attention to the papers [11] and [48, 49].

3. Romanov type problems

This chapter contains an article, which is joint work with Christian Elsholtz and Florian Luca, and which is going to appear in The Ramanujan Journal. Apart from minor changes, mostly in typesetting, the article below is identical with the published version [21].

Romanov type problems

Christian Elsholtz, Florian Luca and Stefan Planitzer

Abstract. Romanov proved that the proportion of positive integers which can be represented as a sum of a prime and a power of 2 is positive. We establish similar results for integers of the form $n = p + 2^{2^k} + m!$ and $n = p + 2^{2^k} + 2^q$ where $m, k \in \mathbb{N}$ and p, q are primes. In the opposite direction Erdős constructed a full arithmetic progression of odd integers none of which is the sum of a prime and a power of two. While we also exhibit in both cases full arithmetic progressions which do not contain any integers of the two forms, respectively, we prove a much better result for the proportion of integers not of these forms:

1. The proportion of positive integers not of the form $p + 2^{2^k} + m!$ is larger than $\frac{3}{4}$.
2. The proportion of positive integers not of the form $p + 2^{2^k} + 2^q$ is at least $\frac{2}{3}$.

3.1. Introduction

An old result of Romanov [66] states that a positive proportion of the positive integers can be written in the form $p + g^k$, where p is a prime and $g \geq 2$ is a positive integer. As there are about $\frac{x}{\log x}$ primes $p \leq x$ and $\left\lfloor \frac{\log x}{\log g} \right\rfloor$ powers $g^k \leq x$ this result implicitly gives some information about the number $r(n)$ of representations of $n = p + g^k$. There are not too many integers $n \leq x$ with a very large number of representations and on average $r(n)$ is bounded. The most prominent special case of Romanov's result is the one concerning sums of primes and powers of 2. Euler [37] observed in a letter to Goldbach that 959 can not be written as the sum of a prime and a power of two. Euler's letter was also mentioned by de Polignac [65] and provides a counter example to a conjecture of de Polignac himself, stating that any odd positive integer is the sum of a prime and a power of 2. In 1950 Erdős [29] and van der Corput [14] independently proved that also the lower density of odd integers not of the form $p + 2^k$ is positive. Here and in the following the lower density of a set $\mathcal{A} \subset \mathbb{N}$ is defined to be

$$\liminf_{x\to\infty} \frac{|\{a \in \mathcal{A} : a \leq x\}|}{x}.$$

Replacing lim inf with lim sup leads to what we call upper density and if lower and upper density coincide we speak of the density of the set $\mathcal{A}$.

Concerning Romanov's theorem one may ask how this result can be generalized. One way would be by replacing the sequence of powers of g with another sequence $(a_n)_{n\geq 1}$. Generalizing a result of Lee [55] who replaced the powers of g by the Fibonacci sequence, Ballot and Luca [4] proved an analogue of Romanov's theorem for the case when $(a_n)_{n\geq 1}$ is a linearly recurrent sequence with certain additional properties. For certain quadratic recurrences $(a_n)_{n\geq 1}$ this was done by Dubickas [16].

We would expect that for many sets $\mathcal{A} \subset \mathbb{N}$, with $|\mathcal{A} \cap [1, x]| \geq c \log x$ for some positive constant c, one can write a positive proportion of integers $n \leq x$ as $n = p + a$, p prime and $a \in \mathcal{A}$. In this paper we study sets $\mathcal{A}$ with $|\mathcal{A} \cap [1, x]| \sim c_{\mathcal{A}} \log x$ but of a quite different nature compared to previous ones. In particular, we study

$$\begin{aligned}
\mathcal{A}_1 &= \{2^{2^k} + m! : k, m \in \mathbb{N}_0\}, \\
\mathcal{A}_2 &= \{2^{2^k} + 2^q : k \in \mathbb{N}_0, q \text{ prime}\}.
\end{aligned}$$

Using the machinery of Romanov [66] we prove the following two theorems.

Theorem 3.1. *The lower density of integers of the form* $p + 2^{2^k} + m!$ *for* $k, m \in \mathbb{N}_0$ *and* p *prime is positive.*

Theorem 3.2. *The lower density of integers of the form* $p + 2^{2^k} + 2^q$ *for* $k \in \mathbb{N}_0$ *and* p, q *prime is positive.*

Concerning integers not of the form $p + 2^{2^k} + m!$ we consider two different questions.

The first one is finding a large set, in the sense of lower density, of odd positive integers not of this form.

The second question is if there is a full arithmetic progression of odd positive integers not of the form $p+2^{2^k}+m!$. The positive answer to this question is given in Theorem 3.4. Note that the density of the set constructed in the proof of Theorem 3.4 is considerably less than the density of the set used in the proof of Theorem 3.3.

Theorem 3.3. *The lower density of odd positive integers not of the form $p+2^{2^k}+m!$ for $k, m \in \mathbb{N}_0$ and p prime is at least $\frac{615850829669273873}{2459565876494606882} > \frac{1}{4}$. The lower density of all positive integers without a representation of the form $p+2^{2^k}+m!$ is therefore larger than $\frac{3}{4}$.*

Theorem 3.4. *There exists a full arithmetic progression of odd positive integers not of the form $p+2^{2^k}+m!$ for $k, m \in \mathbb{N}_0$ and p prime.*

Finally we prove analogous results for integers not of the form $p+2^{2^k}+2^q$.

Theorem 3.5. *There exists a subset of the odd positive integers not of the form $p+2^{2^k}+2^q$, for $k \in \mathbb{N}$ and p, q prime, with lower density $\frac{1}{6}$. The lower density of all positive integers without a representation of the form $p+2^{2^k}+2^q$ is therefore larger than $\frac{2}{3}$.*

Furthermore, there exists a full arithmetic progression of odd positive integers not of the form $p+2^{2^k}+2^q$.

Concerning the last result, we recall that Erdős conjectured that the lower density of the set of positive odd integers not of the form $p+2^k+2^m$ is positive for $k, m \in \mathbb{N}_0$, p prime (see for example [42, Section A19]).

For the proofs of Theorem 3.1 and Theorem 3.2 we apply the method of Romanov [66]. This means that we start with the Cauchy-Schwarz inequality in the form

$$\left(\sum_{\substack{n\le x\\ r_i(n)>0}} 1\right)\left(\sum_{n\le x} r_i(n)^2\right) \ge \left(\sum_{n\le x} r_i(n)\right)^2 \tag{3.1}$$

for $i \in \{1,2\}$, where $r_1(n)$ denotes the number of representations of n in the form $p+2^{2^k}+m!$ and $r_2(n)$ counts the number of representations of n in the form $p+2^{2^k}+2^q$. Note that the first sum on the left hand side of equation (3.1) equals the number of integers less than x having a representation of the required form. It thus suffices to check that

$$\sum_{n\le x} r_i(n) \gg x \text{ and } \sum_{n\le x} r_i(n)^2 \ll x$$

for both $i = 1,\ 2$ in order to get positive lower density for the sets of those integers. The estimates $\sum_{n\le x} r_1(n) \gg x$ and $\sum_{n\le x} r_1(n)^2 \ll x$ are proved in Section 3.3, Lemma 3.11 and Lemma 3.12, respectively. The analogous results for $r_2(n)$ are proved in Section 3.4, Lemma 3.13 and Lemma 3.14, respectively. Theorem 3.3 and Theorem 3.4 are proved at the end of Section 3.3 and Theorem 3.5 at the end of Section 3.4.

3.2. Notation

Let $\mathbb{N}$, as usual, denote the set of positive integers, $\mathbb{N}_0$ the set of non-negative integers and let $\mathbb{P}$ denote the set of primes. The variables p and q will always denote prime numbers. For any prime $p \in \mathbb{P}$ and any positive integer $n \in \mathbb{N}$ let $\nu_p(n)$ denote the p-adic valuation of n, i.e. $\nu_p(n) = k$ where p^k is the highest power of p dividing n. For an integer n, $P(n)$ denotes its largest prime factor. For any set $S \subset \mathbb{N}$ let $S(x) = |S \cap [1, x]|$ denote the counting function of S. As usual φ denotes Euler's totient function and μ the Möbius function. Furthermore, for an odd positive integer n we denote by $t(n)$ the order of 2 mod n. We use the symbols $\ll$, $\gg$, $\mathcal{O}$ and o within the context of the well known Vinogradov and Landau notation.

3.3. Integers of the form $p + 2^{2^k} + m!$

Before proving Lemma 3.11 and Lemma 3.12 we establish and collect several results needed in due course. The following is a classical result due to Legendre (see for example Theorem 2.6.1 and Theorem 2.6.4 in [61]).

Lemma 3.6 (Legendre's formula)**.** *For any prime $p \in \mathbb{P}$ and any positive integer $n \in \mathbb{N}$ we have that*

$$\nu_p(n!) = \sum_{k=1}^{\infty} \left\lfloor \frac{n}{p^k} \right\rfloor .$$

Furthermore, if $\sigma_p(n)$ denotes the sum of base p digits of n, then

$$\nu_p(n!) = \frac{n - \sigma_p(n)}{p-1}.$$

Theorem 3.7. *The equation $2^{x_1} + y_1! = 2^{x_2} + y_2!$ has only four non-negative integer solutions (x_1, y_1, x_2, y_2) with $x_1 > x_2$ where either $x_2 \le 52$ or $y_2 \le 8$. These solutions are*

$$(x_1, y_1, x_2, y_2) \in \{(1,0,0,2), (1,1,0,2), (3,2,2,3), (7,4,5,5)\}.$$

Proof. Suppose that $x_2 \le 52$ and note that $y_1 = 0$ either implies that $y_2 \in \{0, 1\}$ if $x_2 > 0$, which leads to a solution where $x_1 = x_2$, which is excluded, or implies that $x_2 = 0$, whence $x_1 = 1$ and $y_2 = 2$. Hence, the only solution where $y_1 = 0$ is $(x_1, y_1, x_2, y_2) = (1, 0, 0, 2)$. From now on we may suppose that $y_1 \ge 1$. In this case, from Lemma 3.6, we get that $\nu_2(y_1!) \ge \frac{y_1}{2} - 1$. This yields $\frac{y_1}{2} - 1 \le x_2$ and thus $y_1 \le 106$. Since

$$2^{x_2} - y_1! = 2^{x_1} - y_2!,$$

we have $\nu_2(2^{x_2} - y_1!) = \nu_2(2^{x_1} - y_2!)$. Certainly $|2^{x_2} - y_1!| \le 2^{52} + 106!$ which implies that $\nu_2(2^{x_2} - y_1!) \le \frac{\log(2^{52}+106!)}{\log 2} < 816$. If $x_1 \ge 816$ and $y_2 \ge 822$, then $\nu_2(2^{x_1} - y_2!) \ge 816$, a contradiction. The cases where either $x_1 \le 815$ or $y_2 \le 821$ can be checked by a computer search which leads to the solutions

$$(x_1, y_1, x_2, y_2) \in \{(1, 0, 0, 2), (1, 1, 0, 2), (3, 2, 2, 3), (7, 4, 5, 5)\}.$$

Now suppose that $y_2 \le 8$ and consider

$$0 < 2^{x_1} - 2^{x_2} = y_2! - y_1!,$$

which implies that $y_1 \le y_2 \le 8$. In particular, $|y_2! - y_1!| \le 2 \cdot 8!$ and thus

$$\nu_2(y_2! - y_1!) \le \frac{\log(2 \cdot 8!)}{\log 2} < 17.$$

Since $\nu_2(2^{x_1} - 2^{x_2}) = x_2$ we have that $x_2 < 17$ which is included in the case $x_2 \le 52$ treated above. □

Theorem 3.8. *If we exclude solutions arising from interchanging* (x_1, y_1) *and* (x_2, y_2)*, the equation* $2^{x_1} + y_1! = 2^{x_2} + y_2!$ *has only four non-negative integer solutions* (x_1, y_1, x_2, y_2) *with* $(x_1, y_1) \ne (x_2, y_2)$ *and* $(y_1, y_2) \notin \{(1, 0), (0, 1)\}$ *if* $x_1 = x_2$*. These are the solutions presented in Theorem 3.7.*

Proof. We compare the 2-adic and 3-adic valuation of both sides of equivalent forms of the equation $2^{x_1} + y_1! = 2^{x_2} + y_2!$ to get information about the size of the parameters x_1, x_2, y_1 and y_2.

If $x_1 = x_2$ we have that $y_1! = y_2!$ and hence either $y_1 = y_2$ or $(y_1, y_2) \in \{(1, 0), (0, 1)\}$ which leads to the excluded trivial solutions. Therefore, w.l.o.g., we may suppose that $x_1 > x_2$ and write

$$2^{x_2}(2^{x_1 - x_2} - 1) = y_1!((y_1 + 1) \cdots y_2 - 1). \tag{3.2}$$

Next we compute the 2-adic valuation of both sides of the last equality. For the left–hand side we simply have $\nu_2(2^{x_2}(2^{x_1-x_2}-1)) = x_2$ while for the right–hand side we use that the factor $((y_1+1)\cdots y_2 - 1)$ is odd as soon as $y_2 \geq y_1 + 2$ which yields

$$\nu_2(y_1!((y_1+1)\cdots y_2 - 1)) = \begin{cases} \nu_2(y_1!), & \text{if } y_2 \geq y_1 + 2 \\ \nu_2(y_1!) + \nu_2(y_1), & \text{if } y_2 = y_1 + 1. \end{cases}$$

From this, Lemma 3.6 and the fact that $1 \leq \sigma_2(y_1) \leq \frac{\log y_1}{\log 2} + 1$ (note that as in the proof of Theorem 3.7, $y_1 \in \{0, 1\}$ leads to a single non trivial solution listed there), we get the following two inequalities:

$$\begin{aligned} x_2 = \nu_2(2^{x_2}(2^{x_1-x_2}-1)) = \nu_2(y_1!((y_1+1)\cdots y_2 - 1)) &\leq \nu_2(y_1!) + \nu_2(y_1) \\ &< y_1 + \frac{\log y_1}{\log 2} \end{aligned} \tag{3.3}$$

$$\begin{aligned} x_2 = \nu_2(2^{x_2}(2^{x_1-x_2}-1)) = \nu_2(y_1!((y_1+1)\cdots y_2 - 1)) &\geq \nu_2(y_1!) \\ &\geq y_1 - \left(\frac{\log y_1}{\log 2} + 1\right). \end{aligned} \tag{3.4}$$

By Theorem 3.7, we may suppose that $x_2 \geq 5$ without loosing solutions. In this case the last inequality implies $y_1 \leq 2x_2$.

Next we look at

$$2^{x_1} = 2^{x_2} + y_2! - y_1!.$$

Since $2^{x_2} \leq 2^{x_1-1} = \frac{2^{x_1}}{2}$ we have that $y_2! > \frac{2^{x_1}}{2}$, whence we get

$$y_2^{y_2} \geq y_2! > \frac{2^{x_1}}{2},$$

and thus

$$y_2 \log y_2 > (x_1 - 1)\log 2 \text{ and } y_2 > \frac{(x_1-1)\log 2}{\log y_2}.$$

To get the last inequality we used that by Theorem 3.7 we may suppose that $y_2 \geq 9$ whence $\log y_2 > 0$. Now $x_2 \geq 5$ implies that $x_1 \geq 6$. If we would have that $y_2 \leq x_1$ the last inequality would imply that

$$y_2 > \frac{\log 2}{2}\left(\frac{x_1}{\log y_2}\right) > \frac{1}{4}\left(\frac{x_1}{\log x_1}\right). \tag{3.5}$$

In order to prove (3.5) it therefore suffices to prove that $y_2 \leq x_1$ for $x_1 \geq 6$. In order to do so we consider the equation

$$2^{x_1} = y_1!((y_1+1)\cdots y_2 - 1) + 2^{x_2}$$

from which we readily deduce that $y_1! < 2^{x_1}$. This together with $2^{x_1} = y_2! - y_1! + 2^{x_2}$ implies that

$$y_2! < 2 \cdot 2^{x_1}.$$

This implies that $y_2 \leq x_1$, since otherwise $(x_1+1)! \leq 2^{x_1+1}$ which is true for $x_1 \leq 2$ only. By Theorem 3.7 again, we may suppose that $y_2 \geq 9$. In this case, applying Lemma 3.6, we obtain

$$\nu_3(y_2!) \geq \left\lfloor \frac{y_2}{3} \right\rfloor + \left\lfloor \frac{y_2}{9} \right\rfloor \geq \frac{y_2}{3} > \frac{1}{12}\left(\frac{x_1}{\log x_1}\right), \tag{3.6}$$

where the last inequality follows by (3.5). Now we compute the 3-adic valuation of both sides of equation (3.2). By inequality (3.3) and Lemma 3.6 for the right–hand side, we get

$$\begin{aligned} k = \nu_3(y_1!((y_1+1)\cdots y_2 - 1)) \geq \nu_3(y_1!) = \frac{y_1 - \sigma_3(y_1)}{2} &\geq \frac{y_1}{2} - \frac{\log y_1}{\log 3} - 1 \\ &\geq \frac{x_2}{2} - \log(y_1)\left(\frac{1}{2\log 2} + \frac{1}{\log 3}\right) - 1. \end{aligned}$$

Since for the left–hand side of (3.2) we have $3^k | 2^{x_1 - x_2} - 1$, we deduce that $\varphi(3^k) = 2 \cdot 3^{k-1} | x_1 - x_2$. Here we used that 2 is a primitive root modulo any power of 3. This is a direct consequence of Jacobi's observation [50, p. XXXV] that a primitive root modulo p^2 is also a primitive root modulo any higher power of p. Using the above bound for k and the fact that $y_1 \leq 2x_2$, we get

$$x_1 \geq x_2 + 2\cdot 3^{k-1} \geq x_2 + \frac{2}{9} 3^{x_2/2 - \log(y_1)(1/2\log 2 + 1/\log 3)} \geq x_2 + \frac{2 \cdot 3^{x_2/2}}{36 x_2^2} \geq \frac{3^{x_2/2}}{18 x_2^2}. \tag{3.7}$$

Next we find an upper bound for x_1 in terms of x_2. Consider the equation

$$2^{x_1} - y_2! = 2^{x_2} - y_1!.$$

Equation (3.5) yields that $y_2 > \frac{1}{4}\frac{x_1}{\log x_1} > \frac{1}{4}\sqrt{x_1}$. Thus, by Lemma 3.6, $\nu_2(y_2!) > \frac{\sqrt{x_1}}{8} - 1$ and hence $\nu_2(2^{x_1} - y_2!) \geq \frac{\sqrt{x_1}}{8} - 1$.

On the other hand, $|2^{x_2}-y_1!| \leq 2^{x_2}+y_1! \leq 2^{x_2}+(2x_2)^{2x_2} \leq 2\cdot(2x_2)^{2x_2}$. Now $\nu_2(2^{x_2}-y_1!)$ is certainly bounded from above by the highest power of 2 less than $2\cdot(2x_2)^{2x_2}$:

$$2^a \leq 2\cdot(2x_2)^{2x_2} \Leftrightarrow a \leq \frac{2x_2\log(2x_2)}{\log 2}+1.$$

We therefore have that $\nu_2(2^{x_2}-y_1!) \leq 4x_2\log(2x_2)+1$ and putting everything together, we get:

$$\frac{\sqrt{x_1}}{8}-1 \leq \nu_2(2^{x_1}-y_2!) = \nu_2(2^{x_2}-y_1!) \leq 4x_2\log(2x_2)+1,$$

which implies that $x_1 \leq (32x_2\log(2x_2)+16)^2$. Combining this with (3.7), we finally arrive at

$$3^{x_2/2} \leq 18x_2^2(32x_2\log(2x_2)+16)^2.$$

This inequality is valid only for $x_2 \leq 52$ and the solutions satisfying this restriction are given in Theorem 3.7. □

Lemma 3.9. *Let $m_1, m_2, m_3, m_4 \in \mathbb{N}$ such that $m_1 > m_2$, $m_3 > m_4$ and*

$$m_1! - m_2! = m_3! - m_4!. \tag{3.8}$$

Then $(m_1, m_2) = (m_3, m_4)$ or $m_1 = m_3$ and $(m_2, m_4) \in \{(0,1),(1,0)\}$.

Proof. We start with the case where either $m_1 = m_2+1$ or $m_3 = m_4+1$ and w.l.o.g suppose that $m_1 = m_2+1$. If furthermore, $m_2 \leq m_4$, we get from equation (3.8)

$$m_2!\,m_2 = m_4!((m_4+1)\cdots m_3 - 1) \geq m_2!m_4,$$

which leads to $m_2 \geq m_4$ and thus $m_2 = m_4$ which implies $m_1 = m_3$. On the other hand, if $m_1 = m_2+1$ and $m_2 > m_4$ equation (3.8) implies that

$$m_2(m_4+1)\cdots m_2 = (m_4+1)\cdots m_3 - 1, \tag{3.9}$$

and therefore $m_4+1|1$ if $m_3 > m_4+1$ and $m_4+1|m_4$ otherwise, whence $m_4 = 0$ in both cases. Now $m_3 = 1$ implies that $(m_1, m_2) = (1, 0)$ and we are done. Otherwise, if $m_3 \neq 1$, then the right–hand side of (3.9) is odd. In order for the left–hand side to be odd we need $m_2 = 1$, which implies that $m_1 = m_3$.

It remains to consider the case where $m_1 \geq m_2+2$ and $m_3 \geq m_4+2$ and w.l.o.g. we suppose that $m_2 > m_4$. We look at equation (3.8) in the form

$$m_2!((m_2+1)\cdots m_1 - 1) = m_4!((m_4+1)\cdots m_3 - 1). \tag{3.10}$$

By assumption, we have that $\nu_2(m_2!) = \nu_2(m_4!)$ which implies that m_4 is even and $m_2 = m_4 + 1$. We hence may rewrite equation (3.10) to get

$$(m_4+1)\cdots m_1 - m_4 = (m_4+1)\cdots m_3.$$

It follows that $m_4 + 1 | m_4$ which implies that $m_4 = 0$. This leads to $m_2 = 1$ and $m_1 = m_3$. $\square$

Lemma 3.10. *For odd positive n, let $t(n)$ be the order of $2 \bmod n$ and $t(n) = 2^{a(n)}b(n)$ such that $b(n)$ is odd. Then the series*

$$\sum_{\substack{2\nmid n\\ \mu^2(n)=1}} \frac{1}{nt(b(n))}$$

converges.

Proof. Recall that $P(n)$ denotes the largest prime factor of n and observe that if $u|v$ then $t(u)|t(v)$, thus $b(u)|b(v)$ and further $t(b(u))|t(b(v))$. From this and Mertens' formula in the weak form

$$\prod_{p\le x}\left(1+\frac{1}{p}\right) \ll \log x,$$

we get

$$\sum_{\substack{2\nmid n\\ \mu^2(n)=1}} \frac{1}{nt(b(n))} \le \sum_{\substack{p\ge 3\\ p\in\mathbb{P}}} \frac{1}{pt(b(p))} \sum_{\substack{2\nmid m\\ \mu(m)^2=1\\ P(m)<p}} \frac{1}{m} = \sum_{\substack{p\ge 3\\ p\in\mathbb{P}}} \frac{1}{pt(b(p))} \prod_{\substack{q<p\\ q\in\mathbb{P}}}\left(1+\frac{1}{q}\right)$$
$$\ll \sum_{\substack{p\ge 3\\ p\in\mathbb{P}}} \frac{\log p}{pt(b(p))}. \tag{3.11}$$

We split the primes into two subsets $\mathcal{P}$ and $\mathcal{Q}$ and consider the contribution of these sets separately. We set $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{P}_3 \cup \mathcal{P}_4$ where

$$\mathcal{P}_1 := \{p \in \mathbb{P} : t(p) < p^{1/3}\},$$
$$\mathcal{P}_2 := \{p \in \mathbb{P} : P(t(p)) < p^{1/\log\log p}, p \notin \mathcal{P}_1\},$$

$$\mathcal{P}_3 := \{p \in \mathbb{P} : P(t(p)) \in \mathcal{P}_1, p \notin \mathcal{P}_1 \cup \mathcal{P}_2\},$$
$$\mathcal{P}_4 := \{p \in \mathbb{P} : p \leq p_0\},$$

for some fixed p_0 to be chosen later. The set $\mathcal{Q}$ is then defined to be $\mathbb{P}\backslash(\mathcal{P} \cup \{2\})$. We start by showing that

$$\mathcal{P}(x) \ll \frac{x}{(\log x)^3}. \tag{3.12}$$

For $\mathcal{P}_1$, applying an idea of Erdős and Murty [33], we use that $p|2^k - 1$ where $k = t(p)$, whence we have that

$$\prod_{\substack{p \leq x \\ p \in \mathcal{P}_1}} p \quad \Big| \prod_{k \leq x^{1/3}} (2^k - 1).$$

From this, we get

$$2^{\mathcal{P}_1(x)} \leq \prod_{\substack{p \leq x \\ p \in \mathcal{P}_1}} p \leq \prod_{k \leq x^{1/3}} (2^k - 1) \leq 2^{\sum_{k \leq x^{1/3}} k} \leq 2^{x^{2/3}},$$

which shows that

$$\mathcal{P}_1(x) \ll x^{2/3} = o\left(\frac{x}{(\log x)^3}\right). \tag{3.13}$$

To deal with the contribution of the set $\mathcal{P}_2$, we set

$$\Psi(x, y) := |\{n \leq x : P(n) \leq y\}|.$$

By known results on smooth numbers (in particular, a result of Canfield, Erdős and Pomerance from [10, Corollary p. 15]), we have for $y > (\log x)^2$,

$$\Psi(x, y) = \frac{x}{\exp((1 + o(1))u \log u)}, \quad \text{where} \quad u = \frac{\log x}{\log y}, \tag{3.14}$$

as both y and u tend to infinity. For $p \in \mathcal{P}_2$ we may suppose that $p > x^{1/2}$ since there are at most $\mathcal{O}(\pi(x^{1/2})) = \mathcal{O}\big(\frac{x^{1/2}}{\log x}\big) = o\big(\frac{x}{(\log x)^3}\big)$ primes in $\mathcal{P}_2$ less than $\sqrt{x}$. If $p > x^{1/2}$ then $\log \log p > \frac{\log \log x}{2}$ for sufficiently large x and hence for $x^{1/2} < p < x$ in $\mathcal{P}_2$ we have

$$P(t(p)) < p^{1/\log \log p} < x^{2/\log \log x}.$$

Put $y := x^{2/\log \log x}$. Thus, $p - 1$ is a number which is at most x, having a divisor $t(p) > p^{1/3} > x^{1/6}$, whose largest prime factor is at most y. It follows that $p - 1 \leq x$ is a multiple of some number $d > x^{1/6}$ with $P(d) \le y$. For a fixed d, the number of such p is at most $\lfloor \frac{x}{d} \rfloor \le \frac{x}{d}$. Summing over d, we get that

$$\begin{aligned}
\mathcal{P}_2(x) &\ll \sum_{\substack{x^{1/6}<d<x \\ P(d)<y}} \frac{x}{d} = x \int_{x^{1/6}}^{x} \frac{1}{t} d\Psi(t,y) \\
&= x \left(\left(\frac{\Psi(t,y)}{t} \right) \Big|_{t=x^{1/6}}^{t=x} + \int_{x^{1/6}}^{x} \frac{1}{t^2} \Psi(t,y) dt \right) \\
&\ll x \left(\frac{\Psi(x,y)}{x} + \int_{x^{1/6}}^{x} \frac{\Psi(t,y)}{t^2} dt \right).
\end{aligned}$$

Putting $u_0 := \frac{\log x^{1/6}}{\log y} = \frac{1}{12} \log\log x$, we get that $u = \frac{\log t}{\log y} \ge u_0$ for all $t \in [x^{1/6}, x]$, and

$$(1+o(1))u_0 \log u_0 = \left(\frac{1}{12} + o(1) \right) \log\log x \log\log\log x > 4 \log\log x \tag{3.15}$$

for large x. Using (3.14) and (3.15), we thus get that

$$\mathcal{P}_2(x) \ll \frac{x + x\log x}{\exp((1+o(1))u_0 \log u_0)} \ll \frac{x}{(\log x)^3}.$$

Next we consider the contribution of $\mathcal{P}_3$. This set contains primes p such that $p-1$ is divisible by some prime $q > p^{1/\log\log p}$ but $q \in \mathcal{P}_1$. We may assume again that $p > x^{1/2}$, then $q > p^{1/\log\log p} > y^{1/4}$, where as before $y = x^{2/\log\log x}$. Fixing q, the number of primes $p \le x$ such that $p-1$ is a multiple of q is at most $\frac{x}{q}$. Summing up over $q \in \mathcal{P}_1$ and using (3.13) we get that

$$\begin{aligned}
\mathcal{P}_3(x) &\le \sum_{\substack{y^{1/4}<q<x \\ q\in\mathcal{P}_1}} \frac{x}{q} \ll x \int_{y^{1/4}}^{x} \frac{\mathrm{d}\mathcal{P}_1(t)}{t} = x \left(\left(\frac{\mathcal{P}_1(t)}{t} \right) \Big|_{t=y^{1/4}}^{x} + \int_{y^{1/4}}^{x} \frac{\mathcal{P}_1(t)}{t^2} \mathrm{d}t \right) \\
&\ll x \left(\frac{1}{x^{1/3}} + \int_{y^{1/4}}^{x} \frac{\mathrm{d}t}{t^{4/3}} \right) \ll \frac{x}{y^{1/12}} \ll \frac{x}{(\log x)^3}.
\end{aligned}$$

Finally choose p_0 such that for $p > p_0$ we have that $p^{1/3\log\log p} > (\log p)^3$ and get

$$\mathcal{P}_4(x) \ll 1 \ll \frac{x}{(\log x)^3}.$$

We are now ready to prove that the sum on the right hand side of (3.11) converges. For the contribution of primes $p \in \mathcal{P}$ we use the Abel summation formula as well as (3.12) and get

$$\sum_{\substack{p\le x\\ p\in\mathcal{P}}}\frac{\log p}{pt(b(p))}\le\sum_{\substack{p\le x\\ p\in\mathcal{P}}}\frac{\log p}{p}=\int_3^x\frac{\log t}{t}\mathrm{d}\mathcal{P}(t)$$
$$=\frac{\mathcal{P}(t)\log t}{t}\Big|_{t=3}^{x}-\int_3^x\frac{1-\log t}{t^2}\mathcal{P}(t)\mathrm{d}t$$
$$\ll 1+\int_3^x\frac{\log t}{t^2}\frac{t}{(\log t)^3}\mathrm{d}t=1+\int_3^x\frac{\mathrm{d}t}{t(\log t)^2}\ll 1.$$

By the definition of $\mathcal{Q}$ for $p\in\mathcal{Q}$ we have that $q=P(t(p))>p^{1/\log\log p}$ which implies that $q|b(p)$ for large p. Furthermore, $q\notin\mathcal{P}_1$ so $t(q)>q^{1/3}>p^{1/3\log\log p}$. By the choice of the constant p_0 in the definition of $\mathcal{P}_4$ this implies that $t(b(p))\ge t(q)>(\log p)^3$. Finally this implies that

$$\sum_{p\in\mathcal{Q}}\frac{\log p}{pt(b(p))}\le\sum_{n\in\mathbb{N}}\frac{1}{n(\log n)^2}\ll 1,$$

which finishes the proof of the lemma. □

Lemma 3.11. *The following estimate holds:*

$$\sum_{n\le x}r_1(n)\gg x.$$

Proof. We certainly have that

$$\sum_{n\le x}r_1(n)\ge\left(\sum_{p\le x/3}1\right)\left(\sum_{2^{2^k}\le x/3}1\right)\left(\sum_{m!\le x/3}1\right).$$

By the Prime Number Theorem

$$\sum_{p\le x/3}1\sim\frac{x}{3\log\left(\frac{x}{3}\right)}\gg\frac{x}{\log x},\tag{3.16}$$

and $2^{2^k}\le\frac{x}{3}$ implies that $k\le\frac{\log\left(\log\left(\frac{x}{3}\right)\right)-\log 2}{\log 2}$ and hence

$$\sum_{2^{2^k}\le x/3}1\gg\log\log x.\tag{3.17}$$

We use that $m!\le m^m$ and that $m^m\le\frac{x}{3}$ for $m\le\frac{\log x}{2\log\log x}$ and sufficiently large x. This implies that

$$\sum_{m!\le x/3} 1 \gg \frac{\log x}{\log\log x}. \tag{3.18}$$

The bounds in (3.16), (3.17) and (3.18) show that

$$\sum_{n\le x} r_1(n) \gg x.$$

$\square$

Lemma 3.12. *The following estimate holds:*

$$\sum_{n\le x} r_1(n)^2 \ll x.$$

Proof. We begin with the observation that the sum counts exactly the number of solutions of the equation

$$p_1 + 2^{2^{k_1}} + m_1! = p_2 + 2^{2^{k_2}} + m_2!$$

in p_1, p_2, k_1, k_2, m_1 and m_2 where $p_1 + 2^{2^{k_1}} + m_1! \le x$. For fixed k_1, k_2, m_1 and m_2 this amounts to counting pairs of primes (p_1, p_2) such that $p_2 = p_1 + h$, where

$$h := 2^{2^{k_1}} + m_1! - 2^{2^{k_2}} - m_2!.$$

If $h = 0$, then we apply Theorem 3.8 to get that either $(k_1, m_1) = (k_2, m_2)$ or $k_1 = k_2$ and $(m_1, m_2) \in \{(1,0),(0,1)\}$[1]. The number of choices of the form $(p_1, p_2, k_1, k_2, m_1, m_2)$ in this case is

$$\mathcal{O}\left(\frac{x}{\log x}\left(\log\log x \frac{\log x}{\log\log x} + \log\log x\right)\right) = \mathcal{O}(x).$$

If h is odd then one of the primes p_1 and p_2 equals 2 and any choice of (k_1, k_2, m_1, m_2) fixes the other prime. There are

$$\mathcal{O}\left((\log\log x)^2\left(\frac{\log x}{\log\log x}\right)^2\right) = o(x)$$

choices for $(p_1, p_2, k_1, k_2, m_1, m_2)$ in this case. To deal with the remaining even $h \ne 0$ we use a classical sieve bound (cf. for example [63, Theorem 7.3]). In this case, the number

[1]Note that x_1 and x_2 in the non trivial solutions in Theorem 3.8 are never both powers of 2.

of pairs (p_1, p_2) of primes such that $p_2 = p_1 + h$ is

$$\mathcal{O}\left(\frac{x}{(\log x)^2}\prod_{p|h}\left(1+\frac{1}{p}\right)\right).$$

Summing over all choices (k_1, k_2, m_1, m_2) such that $h \neq 0$ is even (this range of summation is indicated by the dash in the superscript of the sum below) we hence need to show that

$$\frac{x}{(\log x)^2}\sum_{(k_1,k_2,m_1,m_2)}{}' \prod_{p|h}\left(1+\frac{1}{p}\right) \ll x. \tag{3.19}$$

Observing that the prime $p = 2$ contributes just a constant factor, this amounts to showing that

$$\sum_{(k_1,k_2,m_1,m_2)}{}' \prod_{\substack{p|h\\p>2}}\left(1+\frac{1}{p}\right) \ll (\log x)^2,$$

which we do in what follows. We now rewrite the left–hand side of the last inequality as

$$\begin{aligned}\sum_{(k_1,k_2,m_1,m_2)}{}' \prod_{\substack{p|h\\p>2}}\left(1+\frac{1}{p}\right) &= \sum_{(k_1,k_2,m_1,m_2)}{}' \sum_{\substack{d|h\\d \text{ odd}}}\frac{\mu(d)^2}{d}\\ &= \sum_{\substack{d \text{ odd}\\ \mu(d)^2=1}}{}' \frac{|\{(k_1,k_2,m_1,m_2) : d|h\}|}{d}.\end{aligned}$$

Therefore we need to study, for a given odd squarefree d, the cardinality of the set

$$S_d := \{(k_1, k_2, m_1, m_2) : d|h, h \neq 0, 2 \nmid h\}.$$

We start with the subset $S_{1,d} \subset S_d$ where

$$S_{1,d} := \{(k_1, k_2, m_1, m_2) \in S_d : m_1 = m_2 \ \text{ or } \ \{m_1, m_2\} = \{0, 1\}\}. \tag{3.20}$$

We thus first deal with

$$\sum_{\substack{d \text{ odd}\\ \mu(d)^2=1}}{}' \frac{|S_{1,d}|}{d}.$$

By (3.20), (m_1, m_2) is chosen in at most $\mathcal{O}\big(\frac{\log x}{\log\log x}\big)$ ways. As for (k_1, k_2), we have $2^{2^{k_1}} \equiv 2^{2^{k_2}} \pmod d$. Since d is odd this implies that $2^{2^{k_1}-2^{k_2}} \equiv 1 \pmod d$. Recall that $t(d)$ is the order of 2 modulo d. The above congruence makes $2^{k_1} \equiv 2^{k_2} \pmod{t(d)}$. As above we write $t(d) = 2^{a(d)}b(d)$, where $b(d)$ is odd and $a(d)$ is some non-negative integer. This implies that $2^{k_1-k_2} \equiv 1 \pmod{b(d)}$. The above cancellation again is justified since $b(d)$ is odd. Hence, for k_2 fixed, k_1 is in a fixed arithmetic progression modulo $t(b(d))$. The number of such k_1 with $2^{2^{k_1}} \leq x$ is of order (up to a constant) at most

$$\left\lfloor \frac{\log\log x}{t(b(d))} \right\rfloor + 1.$$

Since k_2 is chosen in $\mathcal{O}(\log\log x)$ ways we have

$$\begin{aligned}\sideset{}{'}\sum_{\substack{d \text{ odd} \\ \mu(d)^2=1}} \frac{|S_{1,d}|}{d} &\ll \left(\frac{\log x}{\log\log x}\right) \log\log x \Bigg(\log\log x \sum_{\substack{d \text{ odd} \\ \mu(d)^2=1}} \frac{1}{dt(b(d))} + \sum_{\substack{d\leq x \\ d \text{ odd} \\ \mu(d)^2=1}} \frac{1}{d}\Bigg) \\ &\ll (\log x)^2,\end{aligned}$$

where we used Lemma 3.10 and the fact that

$$\sum_{\substack{d\leq x \\ d \text{ odd} \\ \mu(d)^2=1}} \frac{1}{d} \ll \log x.$$

From now on, we deal with $S_d \backslash S_{1,d}$. Any quadruple (k_1, k_2, m_1, m_2) in the above set gives $m_1! - m_2! \neq 0$ and we assume that $m_1 > m_2$. We partition the numbers d in the range of summation into two different sets A and B. We set

$$\begin{aligned}&A := \\ &\left\{ d \in \mathbb{N} \colon \begin{array}{c} 2 \nmid d, \mu(d)^2 = 1, \forall\{(k_1,k_2,m_1,m_2),(k_3,k_4,m_3,m_4)\} \in (S_d\backslash S_{1,d})^2 : \\ 2^{2^{k_1}} + m_1! - 2^{2^{k_2}} - m_2! = 2^{2^{k_3}} + m_3! - 2^{2^{k_4}} - m_4! = h \end{array} \right\}, \\ &B := \\ &\left\{ d \in \mathbb{N} \colon \begin{array}{c} 2 \nmid d, \mu(d)^2 = 1, \exists\{(k_1,k_2,m_1,m_2),(k_3,k_4,m_3,m_4)\} \in (S_d\backslash S_{1,d})^2 : \\ 2^{2^{k_1}} + m_1! - 2^{2^{k_2}} - m_2! \neq 2^{2^{k_3}} + m_3! - 2^{2^{k_4}} - m_4! \end{array} \right\}.\end{aligned}$$

In the set A we thus collect all d for which all solutions in $S_d \backslash S_{1,d}$ give the same h and the set B contains all other d. For $d \in A$ we fix k_1 and k_2 for solutions in $S_d \backslash S_{1,d}$ and get

$$m_1! - m_2! = h - 2^{2^{k_1}} + 2^{2^{k_2}}.$$

The existence of some other element $(k_1, k_2, m_3, m_4) \in S_d \backslash S_{1,d}$ with $m_3 > m_4$ would imply that $m_1! - m_2! = m_3! - m_4!$ which by Lemma 3.9 leads to $(m_1, m_2) = (m_3, m_4)$. Hence, for $d \in A$ and for $(k_1, k_2, m_1, m_2) \in S_d \backslash S_{1,d}$ with $m_1 > m_2$, the last two coordinates are uniquely determined by the first two whence for $d \in A$ we have

$$|(S_d \backslash S_{1,d})| \ll (\log \log x)^2.$$

We thus get that

$$\sum_{d \in A} \frac{|(S_d \backslash S_{1,d})|}{d} \ll (\log \log x)^2 \sum_{d \leq x} \frac{1}{d} \ll (\log x)(\log \log x)^2 = o((\log x)^2).$$

Finally, we deal with the contribution of $d \in B$. By definition we may find two quadruples (k_1, k_2, m_1, m_2) with $m_1 > m_2$ and (k_3, k_4, m_3, m_4) with $m_3 > m_4$ both in $S_d \backslash S_{1,d}$ such that

$$h := 2^{2^{k_1}} + m_1! - 2^{2^{k_2}} - m_2! \neq 2^{2^{k_3}} + m_3! - 2^{2^{k_4}} - m_4! =: h'. \tag{3.21}$$

Let $\mathcal{P}$ be the set of possible prime factors of $d \in B$ which exceed $\log x$. We shall prove that $|\mathcal{P}| = \mathcal{O}((\log x)^5)$. For h, h' in (3.21) we have that they are both divisible by d and thus $d | h - h'$. Every prime factor of d (in particular the ones larger than $\log x$) divides

$$\prod_{k_i, m_i}{}' \left((2^{2^{k_1}} - 2^{2^{k_2}} + m_1! - m_2!) - (2^{2^{k_3}} - 2^{2^{k_4}} + m_3! - m_4!) \right),$$

where the product is taken over all m_i with $m_i! \leq x$ and all k_i with $2^{2^{k_i}} \leq x$ for $i = 1, 2, 3, 4$. The dash indicates that the product is to be taken over the non zero factors only. Since each factor in this product is of size $\mathcal{O}(x)$ any of these factors has at most $\mathcal{O}(\log x)$ prime factors. Furthermore, for the octuple $(k_1, k_2, k_3, k_4, m_1, m_2, m_3, m_4)$ we have $\mathcal{O}((\log \log x)^4 (\frac{\log x}{\log \log x})^4) = \mathcal{O}((\log x)^4)$ choices and altogether we have that

$$|\mathcal{P}| = \mathcal{O}((\log x)^5).$$

Write $d = u_d v_d$, where u_d is divisible by primes $p \leq \log x$ only. Hence the factor v_d is divisible only by primes in $\mathcal{P}$. Then

$$\sum_{d \in B} \frac{|(S_d \backslash S_{1,d})|}{d} \leq \Bigg(\sum_{\substack{u \text{ odd} \\ \mu(u)^2 = 1 \\ P(u) < \log x}} \frac{|(S_u \backslash S_{1,u})|}{u} \Bigg) \Bigg(\sum_{\substack{v \text{ odd} \\ \mu(v)^2 = 1 \\ p | v \Rightarrow p \in \mathcal{P}}} \frac{1}{v} \Bigg),$$

where we used that $S_d \backslash S_{1,d} \subset S_u \backslash S_{1,u}$ if $u \mid d$. For the second sum we have

$$\sum_{\substack{v \text{ odd} \\ \mu(v)^2=1 \\ p|v \Rightarrow p \in \mathcal{P}}} \frac{1}{v} = \prod_{p \in \mathcal{P}} \left(1 + \frac{1}{p}\right) = \mathcal{O}(1),$$

which follows from partial summation and the fact that $\mathcal{P}$ has $\mathcal{O}((\log x)^5)$ elements all larger than $\log x$. It thus remains to bound

$$\sum_{\substack{u \text{ odd} \\ \mu(u)^2=1 \\ P(u)<\log x}} \frac{|(S_u \backslash S_{1,u})|}{u}.$$

For this, we fix (m_1, m_2) with $m_1 > m_2$ not both in $\{0, 1\}$. Then putting $M_{1,2} = m_2! - m_1!$, we need to count the number of (k_1, k_2) such that $2^{2^{k_1}} - 2^{2^{k_2}} \equiv M_{1,2} \pmod{u}$. Analogously as before, for fixed k_2, this puts k_1 into a fixed arithmetic progression modulo $t(b(u))$. The number of k_1 with $2^{2^{k_1}} \leq x$ in this progression is of order $\mathcal{O}\big(\frac{\log\log x}{t(b(u))} + 1\big)$. Thus we have

$$\sum_{\substack{u \text{ odd} \\ \mu(u)^2=1 \\ P(u)<\log x}} \frac{|(S_u \backslash S_{1,u})|}{u} \ll \left(\frac{\log x}{\log\log x}\right)^2 (\log\log x) \times$$
$$\times \left(\log\log x \sum_{\substack{u \text{ odd} \\ \mu(u)^2=1 \\ P(u)<\log x}} \frac{1}{ut(b(u))} + \sum_{\substack{u \text{ odd} \\ \mu(u)^2=1 \\ P(u)<\log x}} \frac{1}{u}\right) \ll (\log x)^2.$$

Here, we used Lemma 3.10 and Mertens' formula, which yields

$$\sum_{\substack{u \text{ odd} \\ \mu(u)^2=1 \\ P(u)<\log x}} \frac{1}{u} = \prod_{3 \leq p \leq \log x} \left(1 + \frac{1}{p}\right) \ll \log\log x.$$

□

Proof of Theorem 3.3. Since the density of integers of the form $p + 2^{2^k} + m!$, $p \in \mathbb{P}$, $m, k \in \mathbb{N}$ and $m < 2^{2^6} - 1$ is zero, we may suppose that $m \geq 2^{2^6} - 1$. In this case we have $m! \equiv 0 \bmod 2^{2^6} - 1$ and for $k \geq 6$ we have that $2^{2^k} \equiv 1 \bmod 2^{2^6} - 1$. If $n \equiv a + 1 \bmod 2^{2^6} - 1$, where a is a residue class $\bmod\, 2^{2^6} - 1$ with $(a, 2^{2^6} - 1) > 1$, then $(n - 2^{2^k} - m!, 2^{2^6} - 1) > 1$ which leaves only finitely many choices for the prime $p = n - 2^{2^k} - m!$. This implies that the proportion of such n with a representation of the form $n = p + 2^{2^k} + m!$ is zero. We have $2^{2^6} - 1 - \varphi(2^{2^6} - 1)$ choices for the residue class a and half of the integers in these residue classes are odd which yields a density of

$$\frac{2^{2^6} - 1 - \varphi(2^{2^6} - 1)}{2 \cdot (2^{2^6} - 1)} = \frac{615850829669273873}{2459565876494606882}.$$

We note that a more refined version of the above argument was used by Habsieger and Roblot [43, Section 3] to prove an upper bound on the proportion of odd integers not of the form $p + 2^k$.

□

Proof of Theorem 3.4. We will show that none of the integers n satisfying the following system of congruences is of the form $p + 2^{2^k} + m!$:

1 mod 2	1 mod 3	3 mod 5
2 mod 7	6 mod 11	3 mod 17
7 mod 19	9 mod 23.	

By the Chinese Remainder Theorem the arithmetic progressions above intersect in a unique arithmetic progression. Let n be an element of this progression and suppose that $n = p + 2^{2^k} + m!$.

If $m \geq 3$ then $n = p + 2^{2^k} + m! \equiv p + 2^{2^k} \bmod 3$. All primes except for 3 are in the residue classes $1, 2 \bmod 3$ and $2^{2^k} \equiv 1 \bmod 3$ for $k \geq 1$. Thus for $m \geq 3$ and $k \geq 1$ we have that $n = p + 2^{2^k} + m! \equiv 1 \bmod 3$ hence the only possible choice for p is $p = 3$.

Next we show that if $p = 3$ then $m < 5$. To do so we use that $2^{2^k} \equiv 1 \bmod 5$ for $k \geq 2$ hence for $m \geq 5$ we are left with $n = 3 + 2^{2^k} + m! \equiv \{0, 2, 4\} \bmod 5$, a contradiction to $n \equiv 3 \bmod 5$.

In the case that $k = 0$ we will show that $m \geq 3$ implies $m < 7$. Let $n = p + 2 + m!$ and $m \geq 3$. Then $n \equiv 1 \bmod 3$ implies that $p \equiv 2 \bmod 3$. If additionally $m \geq 7$, then $n = p + 2 + m! \equiv p + 2 \bmod 7$. Since $n \equiv 2 \bmod 7$, the only possible choice for p is $p = 7$, which contradicts $p \equiv 2 \bmod 3$.

Using the above observations the only cases we need to consider are those of $m = 0$, $m = 1$, $m = 2$, $m = 3, 4$ and $k = 0$ or $p = 3$ and $m = 5, 6$ and $k = 0$.

If $m \in \{0, 1\}$ and we additionally have that p is odd, then $n = p + 2^{2^k} + 1$ is even, a contradiction to $n \equiv 1 \bmod 2$. It remains to deal with the case when $p = 2$. Then we have $n = 2 + 2^{2^k} + 1$ and we get a contradiction from $n \equiv 3 \bmod 5$ which would imply that $2^{2^k} \equiv 0 \bmod 5$.

For the case $m = 2$ we use that $2^{2^k} \equiv 1 \bmod 17$ for $k \geq 3$. Hence for $m = 2$ and $k \geq 3$ we have that $n = p + 2^{2^k} + 2 \equiv p + 3 \bmod 17$ which together with $n \equiv 3 \bmod 17$ leaves us with $p = 17$. We use that $n = 17 + 2^{2^k} + 2 \equiv 2 \bmod 3$ to get a contradiction to $n \equiv 1 \bmod 3$. Since $m = 2$ and $k = 0$ imply $n = p + 4 \equiv p + 1 \bmod 3$ the only possible choice for p in this case is $p = 3$ but $n = 7 \not\equiv 3 \bmod 5$. If $m = 2$ and $k = 1$ then $n = p+6$ and $n \equiv 6 \bmod 11$ implies that $p = 11$. This contradicts $n \equiv 1 \bmod 3$. Last we need to deal with $m = 2$ and $k = 2$. In this case $n = p+18 \equiv p+3 \bmod 5$ and hence $n \equiv 3 \bmod 5$ implies that $p = 5$. Now $n = 23$ does not satisfy the congruence $n \equiv 1 \bmod 3$.

If $m = 3$ and $p = 3$ we have that $n = 9 + 2^{2^k} \equiv \{8, 10, 11, 13\} \bmod 17$ contradicting $n \equiv 3 \bmod 17$. On the other hand, if $m = 3$ and $k = 0$ then $n = p + 8 \equiv p + 3 \bmod 5$ and we get a contradiction as shown above.

For $m = 4$ and $p = 3$ we get $n = 27 + 2^{2^k} \equiv \{9, 11, 12, 14\} \bmod 17$, a contradiction to $n \equiv 3 \bmod 17$. If $m = 4$ and $k = 0$ it follows that $n = p + 26 \equiv p + 7 \bmod 19$ which implies $p = 19$ and $n = 45$. This contradicts $n \equiv 3 \bmod 5$.

In the case when $m = 5$ and $k = 0$ we have that $n = p+122 \equiv p+3 \bmod 17$. Together with $n \equiv 3 \bmod 17$ this only leaves $p = 17$ which contradicts $n \equiv 3 \bmod 5$.

Finally, if $m = 6$ and $k = 0$ then $n = p + 722 \equiv p + 9 \bmod 23$. Together with $n \equiv 9 \bmod 23$ this only leaves $p = 23$ which yields a contradiction to $n \equiv 3 \bmod 5$.

□

3.4. Integers of the form $p + 2^{2^k} + 2^q$

Lemma 3.13. *The following estimate holds:*

$$\sum_{n \leq x} r_2(n) \gg x.$$

Proof. The lemma follows from

$$\sum_{n \leq x} r_2(n) \geq \left(\sum_{\substack{p \leq x/3 \\ p \in \mathbb{P}}} 1 \right) \left(\sum_{2^{2^k} \leq x/3} 1 \right) \left(\sum_{\substack{q \leq \log x/3 \\ q \in \mathbb{P}}} 1 \right).$$

By the Prime Number Theorem we have

$$\sum_{\substack{p\le x/3\\ p\in\mathbb{P}}} 1 \gg \frac{x}{\log x} \qquad \text{and} \qquad \sum_{\substack{q\le \log x/3\\ q\in\mathbb{P}}} 1 \gg \frac{\log x}{\log\log x}.$$

Together with

$$\sum_{2^{2^k}\le x/3} 1 \gg \log\log x$$

this finishes the proof of the lemma. $\square$

Lemma 3.14. *The following estimate holds:*

$$\sum_{n\le x} r_2(n)^2 \ll x.$$

Proof. Again $r_2(n)^2$ counts the number of solutions of the equation

$$p_1 + 2^{2^{k_1}} + 2^{q_1} = p_2 + 2^{2^{k_2}} + 2^{q_2}$$

in p_1, p_2, k_1, k_2, q_1 and q_2 where $p_1 + 2^{2^{k_1}} + 2^{q_1} \le x$. This means counting pairs of primes (p_1, p_2) such that $p_2 = p_1 + h$, where

$$h := 2^{2^{k_1}} + 2^{q_1} - 2^{2^{k_2}} - 2^{q_2}.$$

If $h = 0$ then either $(k_1, q_1) = (k_2, q_2)$ or w.l.o.g. $k_1 > k_2$ and

$$2^{2^{k_2}}\left(2^{2^{k_1}-2^{k_2}} - 1\right) = 2^{q_1}\left(2^{q_2-q_1} - 1\right).$$

Since $2^{2^{k_1}-2^{k_2}} - 1$ and $2^{q_2-q_1} - 1$ are odd we have that $2^{k_2} = q_1$ and hence $k_2 = 1$ and $q_1 = 2$. This leads to $2^{k_1} = q_2$ and hence to $k_1 = 1$ and $q_2 = 2$ a contradiction to $k_1 > k_2$. If $h = 0$ we thus have that $(k_1, q_1) = (k_2, q_2)$ and p_2 is fixed by a choice of p_1, k_1 and q_1. The last three parameters may be chosen in $\mathcal{O}(x)$ ways and we can deal with the contribution of solutions of the equation $p_2 = p_1 + h$ where $h \ne 0$. Since h is even we may directly use the sieve bound from [63, Theorem 7.3] which, after summing over all h, yields an upper bound of order

$$\frac{x}{(\log x)^2} \sum_{(k_1,q_1,k_2,q_2)}{}' \prod_{p\mid h}\left(1 + \frac{1}{p}\right) \tag{3.22}$$

for the sum in the lemma, where the dash indicates that $(k_1, q_1) \neq (k_2, q_2)$. Noting that the contribution of the prime 2 is just a constant factor, we disregard it. Furthermore, $h \leq x$ by definition, and a very crude upper bound for the number of prime factors of h, in particular for those larger than $\log x$, is given by $\frac{\log x}{\log 2}$. We thus get

$$
\begin{aligned}
\sideset{}{'}\sum_{(k_1,q_1,k_2,q_2)} \prod_{\substack{p|h \\ p>2}} \left(1+\frac{1}{p}\right) &\ll \sideset{}{'}\sum_{(k_1,q_1,k_2,q_2)} \underbrace{\left(1+\frac{1}{\log x}\right)^{\log x/\log 2}}_{\leq e^{1/\log 2}} \prod_{\substack{p|h \\ 2<p\leq \log x}} \left(1+\frac{1}{p}\right) \\
&\ll \sideset{}{'}\sum_{(k_1,q_1,k_2,q_2)} \sum_{\substack{d|h \\ d \text{ odd} \\ P(d)\leq \log x}} \frac{\mu(d)^2}{d} \\
&= \sum_{\substack{d\leq x \\ d \text{ odd} \\ P(d)\leq \log x}} \frac{\mu(d)^2}{d} \sideset{}{'}\sum_{\substack{(k_1,q_1,k_2,q_2) \\ d|h}} 1. \qquad (3.23)
\end{aligned}
$$

If we fix k_1, q_1 and k_2, then the fact that $d \mid h$ implies

$$
2^{q_2} \equiv 2^{2^{k_1}} + 2^{q_1} - 2^{2^{k_2}} =: l \bmod d,
$$

where l is a fixed residue class mod d. This puts q_2 in a fixed residue class mod $t(d)$. Since we are counting representations of integers $n \leq x$ we have $q_2 \leq \frac{\log x}{\log 2}$. Hence if $t(d) > \log x$ there are at most two choices for q_2. If $t(d) \leq \log x$ the Brun-Titchmarsh inequality yields an upper bound of

$$
\mathcal{O}\left(\frac{\log x/\log 2}{\varphi(t(d)) \log\left(\log x/t(d)\log 2\right)}\right)
$$

for the number of choices of q_2. We thus get an upper bound of the following order for (3.23)

$$
\log x \log\log x \Bigg(\sum_{\substack{d \text{ odd} \\ P(d)\leq \log x \\ t(d)\leq \log x}} \frac{\mu(d)^2 \left(\log x/\log 2\right)}{d\varphi(t(d)) \log\left(\log x/t(d)\log 2\right)} + \sum_{\substack{d \text{ odd} \\ P(d)\leq \log x \\ t(d)>\log x}} \frac{\mu(d)^2}{d} \Bigg). \qquad (3.24)
$$

As earlier, by Mertens' formula

$$
\sum_{\substack{d \text{ odd} \\ P(d)\leq \log x}} \frac{\mu(d)^2}{d} \ll \log\log x.
$$

To deal with the first sum in (3.24) we use $\varphi(m) \gg \frac{m}{\log\log m}$ (see [68, Theorem 15]) and split the range of summation in two parts and get

$$\sum_{\substack{d \text{ odd} \\ P(d)\le \log x \\ t(d)\le \log x}} \frac{\mu(d)^2 \left({\log x}/{\log 2}\right)}{d\varphi(t(d)) \log\left({\log x}/{t(d)\log 2}\right)} \ll \frac{\log x}{\log\log x} \sum_{\substack{d \text{ odd} \\ P(d)\le \log x \\ t(d)\le \sqrt{\log x}}} \frac{\mu(d)^2 \log\log t(d)}{dt(d)}$$
$$+ (\log x)^{3/4} \sum_{\substack{d \text{ odd} \\ P(d)\le \log x \\ \sqrt{\log x}<t(d)\le \log x}} \frac{\mu(d)^2 \log\log t(d)}{d\sqrt{t(d)}}.$$

By a result of Erdős and Turán [35, 36] the sums

$$\sum_{d \text{ odd}} \frac{\log\log t(d)}{dt(d)} \text{ and } \sum_{d \text{ odd}} \frac{\log\log t(d)}{d\sqrt{t(d)}}$$

converge which altogether proves an upper bound of order $\mathcal{O}((\log x)^2)$ for (3.23) and hence an upper bound of order $\mathcal{O}(x)$ for (3.22). $\square$

Proof of Theorem 3.5. We prove the theorem by showing that the subset of positive integers in the residue class 3 mod 6 having a representation of the form $p + 2^{2^k} + 2^q$ has density 0.

If $k > 0$ then $2^{2^k} = 4^{2^{k-1}}$. The fact that $4^2 \equiv 4 \bmod 6$ puts the term 2^{2^k} into the residue class 4 mod 6 if $k > 0$. Using the same fact again we get for $q = 2l + 1$

$$2^q = 2^{2l+1} = 2 \cdot 4^l \equiv 2 \bmod 6.$$

Furthermore, all primes except 2 and 3 are in the residue classes $\{1, 5\}$ mod 6. Thus if n is in none of the sets

$$\begin{aligned}
S_1 &:= \{p + 2 + 2^q : p, q \in \mathbb{P}\},\\
S_2 &:= \{p + 2^{2^k} + 4 : p \in \mathbb{P}, k \in \mathbb{N}\},\\
S_3 &:= \{2 + 2^{2^k} + 2^q : k \in \mathbb{N}, q \in \mathbb{P}\},\\
S_4 &:= \{3 + 2^{2^k} + 2^q : k \in \mathbb{N}, q \in \mathbb{P}\},
\end{aligned}$$

all of which have density 0, and if n has a representation of the form $n = p + 2^{2^k} + 2^q$, then n is in one of the residue classes

$$\{1,5\} + \{4\} + \{2\} = \{1,5\} \bmod 6.$$

The set

$$S = \{n \in \mathbb{N} : n \equiv 3 \bmod 6\} \backslash (S_1 \cup S_2 \cup S_3 \cup S_4)$$

has density $\frac{1}{6}$, consists of odd integers only and none of its members is of the form $p + 2^{2^k} + 2^q$. This proves the first part of the Theorem.

To find a full arithmetic progression of integers not of the form $p + 2^{2^k} + 2^q$ we will add additional congruences ruling out the integers in the sets S_1, S_2, S_3 and S_4. We claim that none of the integers n satisfying the congruences

3 mod 6	4 mod 5	4 mod 7
9 mod 13	5 mod 17	8 mod 19
20 mod 23	2 mod 29	3 mod 31
10 mod 37		

is of the form $p + 2^{2^k} + 2^q$. By the above considerations, it suffices to check that none of the integers in the sets S_1, S_2, S_3 and S_4 is contained in this arithmetic progression.

We start with the integers in S_1. Take $n = p + 2 + 2^q \in S_1$ and suppose that n is in the arithmetic progression constructed above. We use that except for $q \in \{2,3\}$ we have that $q \equiv \{1,5,7,11\} \bmod 12$ and that for any $l \in \mathbb{N}_0$ we have that

$$2^{12l+1} \equiv 2^{12l+5} \equiv 2 \bmod 5,\ 2^{12l+7} \equiv 2 \bmod 7,\ 2^{12l+11} \equiv 7 \bmod 13.$$

If $q \equiv \{1,5\} \bmod 12$ then $n = p + 2 + 2^q \equiv p + 4 \bmod 5$. Since $n \equiv 4 \bmod 5$ this implies that $p = 5$. Now $7 + 2^{12l+1} \equiv 2 \bmod 7$ and $7 + 2^{12l+5} \equiv 0 \bmod 13$, contradiction to $n \equiv 4 \bmod 7$ and $n \equiv 9 \bmod 13$. In the case of $q = 12l + 7$ we get $n = p + 2 + 2^{12l+7} \equiv p + 4 \bmod 7$ and the only possible choice for p is $p = 7$. Then $9 + 2^{12l+7} \equiv 2 \bmod 5$, a contradiction to $n \equiv 4 \bmod 5$. Finally if $q = 12l + 11$ then $n = p + 2 + 2^{12l+11} \equiv p + 9 \bmod 13$ and from $n \equiv 9 \bmod 13$ we get $p = 13$. Since $n = 15 + 2^{12l+11} \equiv 3 \bmod 5$ we again get a contradiction to $n \equiv 4 \bmod 5$. To finish off the integers in the set S_1 it remains to deal with $q \in \{2,3\}$. If $q = 2$ we have $n = p + 6 \equiv p \bmod 6$. Since $n \equiv 3 \bmod 6$ we are left with $p = 3$ and $n = 9$ which contradicts to $n \equiv 4 \bmod 7$. If $q = 3$ then $n = p + 10$ and from $n \equiv 10 \bmod 37$ we need to have that $p = 37$ and hence $n = 47$. This is impossible since it contradicts to $n \equiv 4 \bmod 5$.

Next we deal with the integers in S_2 and we use that $2^{2^k} \equiv 1 \bmod 17$ for $k \geq 3$. Thus for $k \geq 3$ and $n = p + 2^{2^k} + 4 \in S_2$ we have that $n = p + 2^{2^k} + 4 \equiv p + 5 \bmod 17$. From $n \equiv 5 \bmod 17$ we see that the only admissible choice for p is $p = 17$ and hence $n = 21 + 2^{2^k}$. As above we use that $2^{2^k} \equiv \{2, 4\} \bmod 6$ and thus $21 + 2^{2^k} \equiv \{1, 5\} \bmod 6$ a contradiction to $n \equiv 3 \bmod 6$. We are left with $k \in \{0, 1, 2\}$. For $k = 0$ we get $n = p + 6$ which was ruled out when we dealt with the integers in S_1. If $k = 1$ we have $n = p + 8$ and from $n \equiv 8 \bmod 19$, the only possible choice for p is $p = 19$ and thus $n = 27$. This contradicts to $n \equiv 4 \bmod 5$. Finally, if $k = 2$, we have $n = p + 20$ and from $n \equiv 20 \bmod 23$ we again are left with a single possible choice for p, namely $p = 23$. Now $n = 43$, contradicting to $n \equiv 4 \bmod 5$.

For integers n in the set S_3 we have $n = 2 + 2^{2^k} + 2^q$. If $q = 2$ we have $n \equiv 2^{2^k} \bmod 6$ and again using that $2^{2^k} \in \{2, 4\} \bmod 6$ we get a contradiction to $n \equiv 3 \bmod 6$. If q is odd, then $2^q \equiv 2 \bmod 6$. If, furthermore, $k = 0$ then $n = 4 + 2^q \equiv 0 \bmod 6$ and if $k = 1$ we get $n = 6 + 2^q \equiv 2 \bmod 6$. In both cases this yields a contradiction to $n \equiv 3 \bmod 6$. For $k \geq 2$ and q odd we have that $2^{2^k} \equiv \{16, 24, 25\} \bmod 29$ and $2^q \equiv \{2, 3, 8, 10, 11, 12, 14, 15, 17, 18, 19, 21, 26, 27\} \bmod 29$. For $k \geq 2$ and q odd it is thus true that $2^{2^k} + 2^q \not\equiv 0 \bmod 29$ and thus $n = 2 + 2^{2^k} + 2^q \equiv 2 \bmod 29$ yields a contradiction in this case.

Finally for integers in the set S_4 we apply a similar argument as for integers in the set S_3. For any prime q we have that $2^q \equiv \{1, 2, 4, 8, 16\} \bmod 31$ and for all $k \in \mathbb{N}_0$ we get $2^{2^k} \equiv \{2, 4, 8, 16\} \bmod 31$. Again $2^{2^k} + 2^q \not\equiv 0 \bmod 31$ for any prime q and any non-negative integer k. Thus $n = 3 + 2^{2^k} + 2^q \equiv 3 \bmod 31$ yields a contradiction.

□

3.5. Acknowledgements

The first and the third author are supported by the Austrian Science Fund (FWF): W1230, Doctoral Program ‘Discrete Mathematics’. The second author is supported by grant no. 17-02804S of the Czech Grant Agency. Parts of this research work were done when the second author was visiting the Institute of Analysis and Number Theory at Graz University of Technology and the Max Planck Institute for Mathematics in Bonn, Germany. He wants to thank these institutions and Professor Robert Tichy. Furthermore, we would like to thank the referee for carefully reading the manuscript.

4. Sequences with Property P

This chapter contains an article, which is joint work with Christian Elsholtz, and which appeared in Monatshefte für Mathematik. Apart from a newly added appendix and minor changes, mostly in typesetting, the article below is identical with the published version [22].

On Erdős and Sárközy's sequences with Property P

Christian Elsholtz and Stefan Planitzer

Abstract. A sequence A of positive integers having the property that no element $a_i \in A$ divides the sum $a_j + a_k$ of two larger elements is said to have 'Property P'. We construct an infinite set $S \subset \mathbb{N}$ having Property P with counting function $S(x) \gg \frac{\sqrt{x}}{\sqrt{\log x}(\log\log x)^2(\log\log\log x)^2}$. This improves on an example given by Erdős and Sárközy with a lower bound on the counting function of order $\frac{\sqrt{x}}{\log x}$.

4.1. Introduction

Erdős and Sárközy [34] define a monotonically increasing sequence $A = \{a_1 < a_2 < \ldots\}$ of positive integers to have 'Property P' if $a_i \nmid a_j + a_k$ for $i < j \le k$. They proved that any infinite sequence of integers with Property P has density 0. Schoen [70] showed that if an infinite sequence A has Property P and any two elements in A are coprime then the counting function $A(x) = \sum_{a_i<x} 1$ is bounded from above by $A(x) < 2x^{2/3}$ for infinitely many $x \in \mathbb{N}$ and Baier [3] improved this to $A(x) < (3+\epsilon)x^{2/3}(\log x)^{-1}$ for infinitely many $x \in \mathbb{N}$ and any $\epsilon > 0$. Concerning finite sequences with Property P, Erdős and

Sárközy [34] get the lower bound $\max A(x) \geq \lfloor \frac{x}{3} \rfloor + 1$ by just taking A to be the set $A = \{x, x-1, \ldots, x - \lfloor \frac{x}{3} \rfloor\}$ for $x \in \mathbb{N}$[1].

Erdős and Sárközy also thought about infinite sets with Property P with a large counting function (cf. [34, p. 98]). They observed that the set

$$A = \{q_i^2 : q_i \text{ the } i\text{-th prime with } q_i \equiv 3 \bmod 4\}$$

has Property P. This uses the fact that the square of a prime $p \equiv 3 \bmod 4$ has only the trivial representation $p^2 = p^2 + 0^2$ as the sum of two squares. With this set A they get

$$A(x) \sim \frac{\sqrt{x}}{\log x}.$$

Erdős has asked repeatedly to improve this (see e.g. [30, p. 185], [31, p. 535]) and in particular, Erdős [31, 32] asked if one can do better than $a_n \sim (2n \log n)^2$. He wanted to know if it is possible to have $a_n < n^2$. We will not quite achieve this but we go a considerable step in this direction. First, we observe that a set of squares of integers consisting of precisely k prime factors $p \equiv 3 \bmod 4$ also has Property P. As for any fixed k this would only lead to a moderate improvement, our next idea is to try to choose k increasing with x. In order to do so, we actually use a union of several sets S_i with Property P. Together, this union will have a good counting function throughout all ranges of x. However, in order to ensure that this union of sets with Property P still has Property P, we employ a third idea, namely to equip all members $a \in S_i$ with a special indicator factor. This seems to be the first improvement going well beyond the example given by Erdős and Sárközy since 1970. Our main result will be the following theorem.

Theorem 4.1. *The set $S \subset \mathbb{N}$ constructed explicitly below has Property P and counting function*

$$S(x) \gg \frac{\sqrt{x}}{\sqrt{\log x}(\log\log x)^2(\log\log\log x)^2}.$$

We achieve this improvement by not only considering squares of primes $p \equiv 3 \bmod 4$ but products of squares of such primes. More formally we set

$$S = \bigcup_{i=1}^{\infty} S_i. \tag{4.1}$$

[1] We note that the bound $\max A(x) \geq \lfloor \frac{x}{3} \rfloor + 1$ is true only under the slightly weaker condition that $a_i \nmid a_j + a_k$ for $i < j < k$. In our case, working with $a_i \nmid a_j + a_k$ for $i < j \leq k$, by taking the largest $\lfloor \frac{x}{3} \rfloor$ positive integers less than x we have that $\max A(x) \geq \lfloor \frac{x}{3} \rfloor$.

Here the sets S_i are defined by

$$S_i := \left\{ n \in \mathbb{N} : n = q_i^4 \nu^2 \right\}, \tag{4.2}$$

where ν is the product of exactly i distinct primes $p \equiv 3 \bmod 4$ and we recall that q_i is the i-th prime in the residue class $3 \bmod 4$. The rôle of the q_i is an 'indicator' which uniquely identifies the set S_i a given integer $n \in S$ belongs to. Results from probabilistic number theory like the Theorem of Erdős-Kac suggest that for varying x different sets S_i will yield the main contribution to the counting function $S(x)$. In particular for given $x > 0$ the main contribution comes from the sets S_i with

$$\frac{\log\log\sqrt{x}}{2} - \sqrt{\frac{\log\log\sqrt{x}}{2}} \leq i \leq \frac{\log\log\sqrt{x}}{2} + \sqrt{\frac{\log\log\sqrt{x}}{2}}.$$

The study of sequences with Property P is closely related to the study of primitive sequences, i.e. sequences where no element divides any other and there is a rich literature on this topic (cf. [45, Chapter V]). Indeed a similar idea as the one described above was used by Martin and Pomerance [58] to construct a large primitive set. While Besicovitch [7] proved that there exist infinite primitive sequences with positive upper density, Erdős [26] showed that the lower density of these sequences is always 0. In his proof Erdős used the fact that for a primitive sequence of positive integers the sum $\sum_{i=1}^{\infty} \frac{1}{a_i \log a_i}$ converges. In more recent work Banks and Martin [5] make some progress towards a conjecture of Erdős which states that in the case of a primitive sequence

$$\sum_{i=1}^{\infty} \frac{1}{a_i \log a_i} \leq \sum_{p \in \mathbb{P}} \frac{1}{p \log p}$$

holds. Erdős [27] studied a variant of the Property P problem, also in its multiplicative form.

4.2. Notation

Before we go into details concerning the proof of Theorem 4.1 we need to fix some notation. Throughout this paper $\mathbb{P}$ denotes the set of primes and the letter p (with or without index) will always denote a prime number. We write $\log_k$ for the k-fold iterated logarithm. The functions ω and Ω count, as usual, the prime divisors of a positive integer n without respectively with multiplicity. For two functions $f, g : \mathbb{R} \to \mathbb{R}^+$ the binary relation $f \gg g$ (and analogously $f \ll g$) denotes that there exists a constant $c > 0$ such that for x sufficiently large $f(x) \geq cg(x)$ ($f(x) \leq cg(x)$ respectively). Dependence of the implied constant on certain parameters is indicated by subscripts. The same convention is used for the Landau symbol $\mathcal{O}$ where $f = \mathcal{O}(g)$ is equivalent to $f \ll g$. We write $f = o(g)$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$.

4.3. The set S has Property P

In this section we verify that any union of sets S_i defined in (4.2) has Property P.

Lemma 4.2. *Let n_1, n_2 and n_3 be positive integers. If there exists a prime $p \equiv 3 \bmod 4$ with $p|n_1$ and $p \nmid \gcd(n_2, n_3)$, then*

$$n_1^2 \nmid n_2^2 + n_3^2.$$

Proof. We prove the Lemma by contradiction. Suppose that $n_1^2 | n_2^2 + n_3^2$. By our assumption there exists a prime $p \equiv 3 \bmod 4$ such that $p|n_1$ and $p \nmid \gcd(n_2, n_3)$. Hence, w.l.o.g. $p \nmid n_2$. We have

$$n_2^2 + n_3^2 \equiv 0 \bmod p$$

and since p does not divide n_2, we get that n_2 is invertible mod p. Hence

$$\left(\frac{n_3}{n_2}\right)^2 \equiv -1 \bmod p$$

a contradiction since -1 is a quadratic non-residue mod p. □

Lemma 4.3. *Any union of sets S_i defined in (4.2) has Property P.*

Proof. Suppose by contradiction that there exist $a_i \in S_i$, $a_j \in S_j$ and $a_k \in S_k$ with $a_i < a_j \leq a_k$ and $a_i | a_j + a_k$. First suppose that either $i \neq j$ or $i \neq k$. Define $l \in \{0, 2\}$ to be the largest exponent such that $q_i^l | \gcd(a_i, a_j, a_k)$ where we again recall that q_i was defined as the i-th prime in the residue class 3 mod 4. Then

$$\frac{a_i}{q_i^l} \left| \frac{a_j}{q_i^l} + \frac{a_k}{q_i^l} \right. .$$

By construction of the sets S_i, S_j and S_k we have that $q_i \big| \frac{a_i}{q_i^l}$ and w.l.o.g. $q_i \nmid \frac{a_j}{q_i^l}$. An application of Lemma 4.2 finishes this case.

If $S_i = S_j = S_k$ then $\Omega(a_i) = \Omega(a_j) = \Omega(a_k)$. If there is some prime p with $p | \frac{a_i}{q_i^4}$ and ($p \nmid \frac{a_j}{q_i^4}$ or $p \nmid \frac{a_k}{q_i^4}$) we may again use Lemma 4.2. If no such p exists, then $a_i | a_j$ and

$a_i | a_k$ trivially holds. With the restriction on the number of prime factors we get that $a_i = a_j = a_k$. $\square$

4.4. Products of k distinct primes

In order to establish a lower bound for the counting functions of the sets S_i in (4.2) we need to count square-free integers containing exactly k distinct prime factors $p \equiv 3 \bmod 4$, but no others, where $k \in \mathbb{N}$ is fixed. For $k \geq 2$ and $\pi_k(x) := \#\{n \leq x : \omega(n) = \Omega(n) = k\}$ Landau [52] proved the following asymptotic formula:

$$\pi_k(x) \sim \frac{x(\log_2 x)^{k-1}}{(k-1)! \log x}.$$

We will need a lower bound of similar asymptotic growth as the formula above for the quantity

$$\pi_k(x;4,3) := \#\{n \leq x : p|n \Rightarrow p \equiv 3 \bmod 4,\ \omega(n) = \Omega(n) = k\}.$$

Very recently Meng [60] used tools from analytic number theory to prove a generalization of this result to square-free integers having k prime factors in prescribed residue classes. The following is contained as a special case in [60, Lemma 9] [2]:

Lemma 4.4 (Meng (2016))**.** *For any $A > 0$, uniformly for $2 \leq k \leq A \log\log x$, we have*

$$\pi_k(x;4,3) = \frac{1}{2^k}\frac{x}{\log x}\frac{(\log\log x)^{k-1}}{(k-1)!} \times \\ \left(1 + \frac{k-1}{\log\log x}C(3,4) + \frac{2(k-1)(k-2)}{(\log\log x)^2}h''\left(\frac{2(k-3)}{3\log\log x}\right) + \mathcal{O}_A\left(\frac{k^2}{(\log\log x)^3}\right)\right),$$

where $C(3,4) = \gamma + \sum_{p\in\mathbb{P}}\left(\log\left(1-\frac{1}{p}\right) + \frac{2\lambda(p)}{p}\right)$, γ is the Euler-Mascheroni constant, $\lambda(p)$ is the indicator function of primes in the residue class $3 \bmod 4$ and

$$h(x) = \frac{1}{\Gamma\left(\frac{x}{2}+1\right)}\prod_{p\in\mathbb{P}}\left(1-\frac{1}{p}\right)^{x/2}\left(1+\frac{x\lambda(p)}{p}\right).$$

We will show that Lemma 4.4 with some extra work implies the following Corollary.

[2] We note that in comparison to the arXiv version of [60, Lemma 9], which we used in the original version of this article, the meanwhile published version [59, Lemma 9] slightly changed. Nonetheless, the proof of Corollary 4.5 works with the new version of Meng's result with only minimal modifications and we present a modified proof in an appendix to this chapter below. The presentation of the alternative proof in the appendix is such that it may be read independently of the proof given here.

Corollary 4.5. *Uniformly for* $\frac{\log\log x}{2} - 1 \leq k \leq \frac{\log\log x}{2} + \sqrt{\frac{\log\log x}{2}}$ *we have*

$$\pi_k(x;4,3) \gg \frac{1}{2^k}\frac{x}{\log x}\frac{(\log_2 x)^{k-1}}{(k-1)!}.$$

Proof. In view of Lemma 4.4 and with $k \sim \frac{\log\log x}{2}$ we see that it suffices to check that, independent of the choice of k and for sufficiently large x, there exists a constant $c > 0$ such that

$$1 + \frac{C(3,4)}{2} + \frac{1}{2}h''\left(\frac{2(k-3)}{3\log\log x}\right) \geq c. \tag{4.3}$$

Note that the left hand side of the above inequality is exactly the coefficient of the main term $\frac{1}{2^k}\frac{x}{\log x}\frac{(\log_2 x)^{k-1}}{(k-1)!}$ for k in the range given in the Corollary. The constant $C(3,4)$ does not depend on k. Using Mertens' Formula (cf. [74, p. 19: Theorem 1.12]) in the form

$$\sum_{\substack{p\in\mathbb{P}\\ p\leq x}} \log\left(1-\frac{1}{p}\right) = -\gamma - \log\log x + o(1)$$

we get

$$C(3,4) = \gamma + \sum_{p\in\mathbb{P}}\left(\log\left(1-\frac{1}{p}\right) + \frac{2\lambda(p)}{p}\right) = 2M(3,4),$$

where $M(3,4)$ is the constant appearing in

$$\sum_{p\in\mathbb{P}}\frac{\lambda(p)}{p} = \frac{\log\log x}{2} + M(3,4) + \mathcal{O}\left(\frac{1}{\log x}\right),$$

which was studied by Languasco and Zaccagnini in [53][3]. The computational results of Languasco and Zaccagnini imply that $0.0482 < M(3,4) < 0.0483$ and hence allow for the following lower bound for $C(3,4)$:

$$C(3,4) = 2M(3,4) > 0.0964. \tag{4.4}$$

It remains to get a lower bound for $h''\left(\frac{2(k-3)}{3\log\log x}\right)$, where the function h is defined as in Lemma 4.4. A straight forward calculation yields that

$$h' = \prod_{p\in\mathbb{P}}\left(1-\frac{1}{p}\right)^{x/2} \times$$

[3]Note that our constant $M(3,4)$ corresponds to the constant $M(4,3)$ in the work of Languasco and Zaccagnini.

$$\left(1+\frac{x\lambda(p)}{p}\right)\frac{\Gamma\left(\frac{x}{2}+1\right)\left(\sum_{p\in\mathbb{P}}\frac{1}{2}\log\left(1-\frac{1}{p}\right)+\frac{\lambda(p)}{p+x\lambda(p)}\right)-\frac{1}{2}\Gamma'\left(\frac{x}{2}+1\right)}{\Gamma\left(\frac{x}{2}+1\right)^2}$$

and

$$h''(x)=f(x)\prod_{p\in\mathbb{P}}\left(1-\frac{1}{p}\right)^{x/2}\left(1+\frac{x\lambda(p)}{p}\right),$$

where

$$\begin{aligned}f(x)=&\frac{\left(\sum_{p\in\mathbb{P}}\frac{1}{2}\log\left(1-\frac{1}{p}\right)+\frac{\lambda(p)}{p+x\lambda(p)}\right)^2}{\Gamma\left(\frac{x}{2}+1\right)}-\frac{\Gamma''\left(\frac{x}{2}+1\right)}{4\Gamma\left(\frac{x}{2}+1\right)^2}-\frac{\sum_{p\in\mathbb{P}}\frac{\lambda(p)}{(p+\lambda(p)x)^2}}{\Gamma\left(\frac{x}{2}+1\right)}\\&-\frac{\Gamma'\left(\frac{x}{2}+1\right)\left(\sum_{p\in\mathbb{P}}\frac{1}{2}\log\left(1-\frac{1}{p}\right)+\frac{\lambda(p)}{p+x\lambda(p)}\right)}{\Gamma\left(\frac{x}{2}+1\right)^2}+\frac{\Gamma'\left(\frac{x}{2}+1\right)^2}{2\Gamma\left(\frac{x}{2}+1\right)^3}.\end{aligned}$$

Note that for $x\to\infty$ and $\frac{\log\log x}{2}-1\leq k\leq\frac{\log\log x}{2}+\sqrt{\frac{\log\log x}{2}}$ the term $\frac{2(k-3)}{3\log\log x}$ gets arbitrarily close to $\frac{1}{3}$. Hence we may suppose that $\frac{99}{300}\leq\frac{2(k-3)}{3\log\log x}\leq\frac{101}{300}$ and it suffices to find a lower bound for $h''(x)$ where $\frac{99}{300}\leq x\leq\frac{101}{300}$. For x in this range Mathematica provides the following bounds on the Gamma function and its derivatives

$$\begin{aligned}0.9271\leq\Gamma\left(\frac{x}{2}+1\right)&\leq 0.9283\\-0.3104\leq\Gamma'\left(\frac{x}{2}+1\right)&\leq -0.3058\\1.3209\leq\Gamma''\left(\frac{x}{2}+1\right)&\leq 1.3302.\end{aligned}$$

Furthermore, we have

$$\sum_{p\in\mathbb{P}}\frac{\lambda(p)}{(p+x)^2}<\sum_{p\in\mathbb{P}}\frac{\lambda(p)}{p^2}<\sum_{\substack{p\in\mathbb{P}\\p\leq 10^4}}\frac{\lambda(p)}{p^2}+\sum_{n>10^4}\frac{1}{n^2}<0.1485+\int_{x=10^4}^{\infty}\frac{\mathrm{d}x}{x^2}=0.1486.$$

Later we will use that

$$\begin{aligned}\sum_{p\in\mathbb{P}}\left(\frac{1}{2}\log\left(1-\frac{1}{p}\right)+\frac{\lambda(p)}{p+x}\right)&=\sum_{p\in\mathbb{P}}\left(\frac{1}{2}\log\left(1-\frac{1}{p}\right)+\frac{\lambda(p)}{p}\right)-x\sum_{p\in\mathbb{P}}\frac{\lambda(p)}{p^2+px}\\&>\sum_{p\in\mathbb{P}}\left(\frac{1}{2}\log\left(1-\frac{1}{p}\right)+\frac{\lambda(p)}{p}\right)-x\sum_{p\in\mathbb{P}}\frac{\lambda(p)}{p^2}\\&=-\frac{\gamma}{2}+M(3,4)-x\sum_{p\in\mathbb{P}}\frac{\lambda(p)}{p^2}>-0.2905,\end{aligned}$$

and

$$\sum_{p\in\mathbb{P}}\left(\frac{1}{2}\log\left(1-\frac{1}{p}\right)+\frac{\lambda(p)}{p+x}\right)<\sum_{p\in\mathbb{P}}\left(\frac{1}{2}\log\left(1-\frac{1}{p}\right)+\frac{\lambda(p)}{p}\right)=-\frac{\gamma}{2}+M(3,4)$$
$$<-0.2403.$$

Finally, using $\log(1+\frac{x}{p})\le\frac{x}{p}$, we get

$$\begin{aligned}0\le\prod_{p\in\mathbb{P}}\left(1-\frac{1}{p}\right)^{x/2}\left(1+\frac{x\lambda(p)}{p}\right)&\le\exp\left(x\left(\sum_{p\in\mathbb{P}}\left(\frac{1}{2}\log\left(1-\frac{1}{p}\right)+\frac{\lambda(p)}{p}\right)\right)\right)\\&=\exp\left(x\left(-\frac{\gamma}{2}+M(3,4)\right)\right)<\exp\left(-\frac{99}{300}\cdot 0.2403\right)<0.9238.\end{aligned}$$

Applying the explicit bounds calculated above, for $\frac{99}{300}\le x\le\frac{101}{300}$ we obtain:

$$f(x)\ge\frac{0.2403^2}{0.9283}-\frac{1.3302}{4\cdot 0.9271^2}-\frac{0.1486}{0.9271}-\frac{0.3104\cdot 0.2905}{0.9271^2}+\frac{0.3058^2}{2\cdot 0.9283^3}>-0.5315.$$

This implies for sufficiently large x:

$$h''\left(\frac{2(k-3)}{3\log\log x}\right)>-0.492.$$

Together with (4.4) this leads to an admissible choice of $c=0.802$ in (4.3). □

4.5. The counting function $S(x)$

Proof of Theorem 4.1. As in (4.1) we set

$$S=\bigcup_{i=1}^{\infty}S_i$$

where the sets S_i are defined as in (4.2). The set S has Property P by Lemma 4.3 and it remains to work out a lower bound for the size of the counting function $S(x)$. For sufficiently large x there exists a uniquely determined integer $k\in\mathbb{N}$ such that $e^{2e^{2k}}\le x<e^{2e^{2(k+1)}}$ hence

$$k\le\frac{\log_2\sqrt{x}}{2}<k+1. \tag{4.5}$$

It depends on the size of x, which S_i makes the largest contribution. For a given x we take several sets $S_{k+2},S_{k+3},\dots,S_{k+l}$, $l=\lfloor\sqrt{\frac{\log_2\sqrt{x}}{2}}\rfloor$, as the number of prime factors

$p \equiv 3 \bmod 4$ of a typical integer less than x is in

$$\left[\frac{\log_2 x}{2} - \sqrt{\frac{\log_2 x}{2}}, \frac{\log_2 x}{2} + \sqrt{\frac{\log_2 x}{2}}\right].$$

Using Corollary 4.5 as well as the fact that the i-th prime in the residue class 3 mod 4 is asymptotically of size $2i \log i$ for given $2 \le j \le l$ we get

$$S_{k+j}(x) \gg \underbrace{\frac{\sqrt{\frac{x}{16(k+j)^4 \log^4(k+j)}}}{\log\left(\sqrt{\frac{x}{16(k+j)^4 \log^4(k+j)}}\right)}}_{\mathrm{F}_1} \cdot \underbrace{\frac{\left(\log_2 \sqrt{\frac{x}{16(k+j)^4 \log^4(k+j)}}\right)^{k+j-1}}{2^{k+j}(k+j-1)!}}_{\mathrm{F}_2}. \tag{4.6}$$

We deal with the fractions F_1 and F_2 on the right hand side of (4.6) separately. With the given range of j and (4.5) we have that

$$\mathrm{F}_1 \gg \frac{\sqrt{x}}{\log x (\log_2 x)^2 (\log_3 x)^2}.$$

It remains to deal with F_2. Using the given range of k and j we have that $k+j \le \log_2 \sqrt{x}$ and, again for sufficiently large x, for the numerator of F_2 we get

$$\begin{aligned}
\log_2^{k+j-1} \frac{\sqrt{x}}{4(k+j)^2 \log^2(k+j)} &\gg (\log(\log \sqrt{x} - \log 4 - 2\log_3 \sqrt{x} - 2\log_4 \sqrt{x}))^{k+j-1} \\
&\gg (\log(\log \sqrt{x} - 5\log_3 \sqrt{x}))^{k+j-1} \\
&= \left(\log_2 \sqrt{x} + \log\left(1 - \frac{5\log_3 \sqrt{x}}{\log \sqrt{x}}\right)\right)^{k+j-1} \\
&\gg \left(\log_2 \sqrt{x} - \frac{10\log_3 \sqrt{x}}{\log \sqrt{x}}\right)^{k+j-1} \\
&\gg \left(1 - \frac{10\log_3 \sqrt{x}}{\log \sqrt{x} \log_2 \sqrt{x}}\right)^{\frac{\log_2 \sqrt{x}}{2} + \sqrt{\frac{\log_2 \sqrt{x}}{2}} - 1} \log_2^{k+j-1} \sqrt{x} \\
&\gg \log_2^{k+j-1} \sqrt{x}.
\end{aligned}$$

Here we used that

$$\lim_{x\to\infty} \left(1 - \frac{10\log_3 \sqrt{x}}{\log \sqrt{x} \log_2 \sqrt{x}}\right)^{\frac{\log_2 \sqrt{x}}{2} + \sqrt{\frac{\log_2 \sqrt{x}}{2}} - 1} = 1$$

and that for $0 \le y \le \frac{1}{2}$ we certainly have that $\log(1-y) \ge -2y$. To deal with the denominator of F_2 we apply Stirling's Formula and get

$$\begin{aligned}(k+j-1)! &\ll \left(\frac{k+j-1}{e}\right)^{k+j-1}\sqrt{k+j-1} \ll \left(\frac{\log_2\sqrt{x}+2(j-1)}{2e}\right)^{k+j-1}\sqrt{\log_2 x}\\ &\ll (\log_2\sqrt{x}+2(j-1))^{k+j-1}\frac{\sqrt{\log_2 x}}{2^{k+j-1}e^{\frac{\log_2\sqrt{x}}{2}+j-2}}\\ &\ll (\log_2\sqrt{x}+2(j-1))^{k+j-1}\frac{\sqrt{\log_2 x}}{2^{k+j-1}e^{j-2}\sqrt{\log x}}.\end{aligned}$$

Altogether we get

$$\begin{aligned}\mathrm{F}_2 &\gg \frac{\sqrt{\log x}}{\sqrt{\log_2 x}}e^{j-2}\left(\frac{\log_2\sqrt{x}}{\log_2\sqrt{x}+2(j-1)}\right)^{k+j-1}\\ &\gg \frac{\sqrt{\log x}}{\sqrt{\log_2 x}}e^{j-2}\left(\frac{\log_2\sqrt{x}}{\log_2\sqrt{x}+2(j-1)}\right)^{\frac{\log_2\sqrt{x}}{2}+j-1}.\end{aligned} \tag{4.7}$$

Since

$$\left(\frac{\log_2\sqrt{x}}{\log_2\sqrt{x}+2(j-1)}\right)^{\frac{\log_2\sqrt{x}}{2}} \sim \frac{1}{e^{j-1}}$$

it suffices to check that for any $x > 0$ and for our choices of j there exists a fixed constant $c > 0$ such that

$$\left(1+\frac{2(j-1)}{\log_2\sqrt{x}}\right)^{1-j} \ge c. \tag{4.8}$$

For $j \ge 2$ we have that $\left(1+\frac{2(j-1)}{\log_2\sqrt{x}}\right)^{1-j}$ is monotonically decreasing in j and get

$$\left(1+\frac{2(j-1)}{\log_2\sqrt{x}}\right)^{1-j} \ge \left(1+\frac{2\sqrt{\frac{\log_2\sqrt{x}}{2}}}{\log_2\sqrt{x}}\right)^{-\sqrt{\frac{\log_2\sqrt{x}}{2}}} = \left(1+\frac{1}{\sqrt{\frac{\log_2\sqrt{x}}{2}}}\right)^{-\sqrt{\frac{\log_2\sqrt{x}}{2}}} \ge \frac{1}{e}.$$

Therefore, for $j \ge 2$ the constant c in (4.8) may be chosen as $c = \frac{1}{e}$ for sufficiently large x. Together with (4.7) this implies

$$F_2 \gg \frac{\sqrt{\log x}}{\sqrt{\log_2 x}}.$$

Altogether for the counting function of any of the sets S_i with $\lfloor \frac{\log_2 \sqrt{x}}{2} \rfloor + 2 \le i \le \lfloor \frac{\log_2 \sqrt{x}}{2} \rfloor + \lfloor \sqrt{\frac{\log_2 \sqrt{x}}{2}} \rfloor$ we have

$$S_i(x) \gg \frac{\sqrt{x}}{\sqrt{\log x}(\log_2 x)^{5/2}(\log_3 x)^2}.$$

Summing these contributions up we finally get

$$S(x) \gg \frac{\sqrt{x}}{\sqrt{\log x}(\log_2 x)^2(\log_3 x)^2}.$$

□

4.6. Acknowledgements

The authors are supported by the Austrian Science Fund (FWF): W1230, Doctoral Program 'Discrete Mathematics'. Parts of this research work were done when the first author was visiting the FIM at ETH Zürich, and the second author was visiting the Institut Élie Cartan de Lorraine of the University of Lorraine. The authors thank these institutions for their hospitality. The authors are also grateful to the referee for suggestions on the manuscript.

Appendix

In proving Corollary 4.5 we used a result from the 2016 arXiv version of Meng's [60] paper on large bias for integers with prime factors in arithmetic progressions. Meanwhile this paper was published, and the statement of the result we used was slightly modified compared to the arXiv version. For completeness' sake, by just modifying the calculations and leaving most of the remaining text unchanged, we adjust our proof of Corollary 4.5 to the following special case of [59, Lemma 9].

Lemma 4.6 (Meng (2018))**.** *For any $A > 0$, uniformly for $2 \le k \le A \log\log x$, we have*

$$\pi_k(x;4,3) = \frac{1}{2^k}\frac{x}{\log x}\frac{(\log\log x)^{k-1}}{(k-1)!}\times \left(1 + \frac{k-1}{\log\log x}C(3,4) + \frac{4(k-1)(k-2)}{(\log\log x)^2}\tilde{h}\left(\frac{2(k-3)}{\log\log x}\right) + \mathcal{O}_A\left(\frac{k^3}{(\log\log x)^4}\right)\right),$$

where $C(3,4) = \gamma + \sum_{p\in\mathbb{P}}\left(\log\left(1-\frac{1}{p}\right) + \frac{2\lambda(p)}{p}\right)$, γ is the Euler-Mascheroni constant,

$\lambda(p)$ is the indicator function of primes in the residue class $3 \bmod 4$ *and*

$$\tilde{h}(x) = \int_0^1 h''(tx)(1-t)\mathrm{d}t,$$

where

$$h(x) = \frac{1}{\Gamma\left(\frac{x}{2}+1\right)} \prod_{p\in\mathbb{P}} \left(1-\frac{1}{p}\right)^{x/2} \left(1+\frac{x\lambda(p)}{p}\right).$$

proof Of Corollary 4.5. In view of Lemma 4.4 and with $k \sim \frac{\log\log x}{2}$ we see that it suffices to check that, independent of the choice of k and for sufficiently large x, there exists a constant $c > 0$ such that

$$1 + \frac{C(3,4)}{2} + \tilde{h}\left(\frac{2(k-3)}{\log\log x}\right) \geq c. \tag{4.9}$$

Note that the left hand side of the above inequality is exactly the coefficient of the main term $\frac{1}{2^k}\frac{x}{\log x}\frac{(\log_2 x)^{k-1}}{(k-1)!}$ for k in the range given in the Corollary. The constant $C(3,4)$ does not depend on k. Using Mertens' Formula (cf. [74, p. 19: Theorem 1.12]) in the form

$$\sum_{\substack{p\in\mathbb{P}\\ p\leq x}} \log\left(1-\frac{1}{p}\right) = -\gamma - \log\log x + o(1)$$

we get

$$C(3,4) = \gamma + \sum_{p\in\mathbb{P}} \left(\log\left(1-\frac{1}{p}\right) + \frac{2\lambda(p)}{p}\right) = 2M(3,4),$$

where $M(3,4)$ is the constant appearing in

$$\sum_{p\in\mathbb{P}} \frac{\lambda(p)}{p} = \frac{\log\log x}{2} + M(3,4) + \mathcal{O}\left(\frac{1}{\log x}\right),$$

which was studied by Languasco and Zaccagnini in [53][4]. The computational results of Languasco and Zaccagnini imply that $0.0482 < M(3,4) < 0.0483$ and hence allow for the following lower bound for $C(3,4)$:

$$C(3,4) = 2M(3,4) > 0.0964. \tag{4.10}$$

It remains to get a lower bound for $\tilde{h}\left(\frac{2(k-3)}{\log\log x}\right)$, where the function $\tilde{h}$ is defined as in

[4]Note that our constant $M(3,4)$ corresponds to the constant $M(4,3)$ in the work of Languasco and Zaccagnini.

Lemma 4.4. First we observe that for $x \to \infty$ and $\frac{\log\log x}{2} - 1 \le k \le \frac{\log\log x}{2} + \sqrt{\frac{\log\log x}{2}}$ the term $\frac{2(k-3)}{\log\log x}$ gets arbitrarily close to 1. Hence we may suppose that $0.999 \le \frac{2(k-3)}{\log\log x} \le 1.001$ and it suffices to find a lower bound for $\tilde{h}(x)$ where $0.999 \le x \le 1.001$. One possible choice for a lower bound in this case is certainly given by

$$-\max_{0.999 \le x \le 1.001} |\tilde{h}(x)|.$$

Since

$$\begin{aligned} |\tilde{h}(x)| = \left| \int_0^1 h''(tx)(1-t)\mathrm{d}t \right| &\le \int_0^{\frac{1}{2}} |h''(tx)||(1-t)|\mathrm{d}t + \int_{\frac{1}{2}}^1 |h''(tx)||(1-t)|\mathrm{d}t \\ &\le \frac{1}{2}\left(\max_{0\le t\le \frac{1}{2}} |h''(tx)| + \frac{1}{2} \max_{\frac{1}{2}\le t\le 1} |h''(tx)| \right), \end{aligned}$$

and with $0.999 \le x \le 1.001$ this reduces to find upper bounds for $\max_{0\le y\le 0.5005} |h''(y)|$ and $\max_{0.4995\le y\le 1.001} |h''(y)|$. A straight forward calculation yields that

$$\begin{aligned} h'(y) = &\prod_{p\in\mathbb{P}} \left(1-\frac{1}{p}\right)^{y/2} \times \\ &\left(1+\frac{y\lambda(p)}{p}\right) \frac{\Gamma\left(\frac{y}{2}+1\right)\left(\sum_{p\in\mathbb{P}} \frac{1}{2}\log\left(1-\frac{1}{p}\right) + \frac{\lambda(p)}{p+y\lambda(p)}\right) - \frac{1}{2}\Gamma'\left(\frac{y}{2}+1\right)}{\Gamma\left(\frac{y}{2}+1\right)^2} \end{aligned}$$

and

$$h''(y) = f(y)\prod_{p\in\mathbb{P}} \left(1-\frac{1}{p}\right)^{y/2}\left(1+\frac{y\lambda(p)}{p}\right),$$

where

$$\begin{aligned} f(y) = &\frac{\left(\sum_{p\in\mathbb{P}} \frac{1}{2}\log\left(1-\frac{1}{p}\right) + \frac{\lambda(p)}{p+y\lambda(p)}\right)^2}{\Gamma\left(\frac{y}{2}+1\right)} - \frac{\Gamma''\left(\frac{y}{2}+1\right)}{4\Gamma\left(\frac{y}{2}+1\right)^2} - \frac{\sum_{p\in\mathbb{P}} \frac{\lambda(p)}{(p+\lambda(p)y)^2}}{\Gamma\left(\frac{y}{2}+1\right)} \\ &- \frac{\Gamma'\left(\frac{y}{2}+1\right)\left(\sum_{p\in\mathbb{P}} \frac{1}{2}\log\left(1-\frac{1}{p}\right) + \frac{\lambda(p)}{p+y\lambda(p)}\right)}{\Gamma\left(\frac{y}{2}+1\right)^2} + \frac{\Gamma'\left(\frac{y}{2}+1\right)^2}{2\Gamma\left(\frac{y}{2}+1\right)^3}. \end{aligned} \tag{4.11}$$

Bounds for the Gamma function and its derivatives for y in the ranges indicated above can be found in Table 4.1. Furthermore, we have

	$0 \leq y \leq 0.5005$		$0.4995 \leq y \leq 0.1001$	
function	lower bound	upper bound	lower bound	upper bound
$\left\lvert\Gamma\left(\frac{y}{2}+1\right)\right\rvert$	0.9063	1	0.8856	0.9065
$\left\lvert\Gamma'\left(\frac{y}{2}+1\right)\right\rvert$	0.2058	0.5773	0.0001	0.2065
$\left\lvert\Gamma''\left(\frac{y}{2}+1\right)\right\rvert$	1.1316	1.9782	0.8293	1.1327

Table 4.1.: Bounds for the Gamma function and its derivatives, computed with Mathematica.

$$\sum_{p\in\mathbb{P}} \frac{\lambda(p)}{(p+y)^2} < \sum_{p\in\mathbb{P}} \frac{\lambda(p)}{p^2} < \sum_{\substack{p\in\mathbb{P}\\ p\leq 10^4}} \frac{\lambda(p)}{p^2} + \sum_{n>10^4} \frac{1}{n^2} < 0.1485 + \int_{t=10^4}^{\infty} \frac{\mathrm{d}t}{t^2} = 0.1486.$$

Later we will use that

$$\begin{aligned}
\sum_{p\in\mathbb{P}} \left(\frac{1}{2}\log\left(1-\frac{1}{p}\right)+\frac{\lambda(p)}{p+y}\right) &= \sum_{p\in\mathbb{P}} \left(\frac{1}{2}\log\left(1-\frac{1}{p}\right)+\frac{\lambda(p)}{p}\right) - y\sum_{p\in\mathbb{P}} \frac{\lambda(p)}{p^2+py} \\
&> \sum_{p\in\mathbb{P}} \left(\frac{1}{2}\log\left(1-\frac{1}{p}\right)+\frac{\lambda(p)}{p}\right) - y\sum_{p\in\mathbb{P}} \frac{\lambda(p)}{p^2} \\
&= -\frac{\gamma}{2} + M(3,4) - y\sum_{p\in\mathbb{P}} \frac{\lambda(p)}{p^2} \\
&> \begin{cases} -0.3148, & \text{if } 0 \leq y \leq 0.5005 \\ -0.3892, & \text{if } 0.4995 \leq y \leq 1.001, \end{cases}
\end{aligned}$$

and

$$\begin{aligned}
\sum_{p\in\mathbb{P}} \left(\frac{1}{2}\log\left(1-\frac{1}{p}\right)+\frac{\lambda(p)}{p+y}\right) < \sum_{p\in\mathbb{P}} \left(\frac{1}{2}\log\left(1-\frac{1}{p}\right)+\frac{\lambda(p)}{p}\right) &= -\frac{\gamma}{2} + M(3,4) \\
&< -0.2403.
\end{aligned}$$

Finally, using $\log(1+\frac{y}{p}) \leq \frac{y}{p}$, we get

$$\begin{aligned}
0 &\leq \prod_{p\in\mathbb{P}} \left(1-\frac{1}{p}\right)^{y/2} \left(1+\frac{y\lambda(p)}{p}\right) \leq \exp\left(y\left(\sum_{p\in\mathbb{P}} \left(\frac{1}{2}\log\left(1-\frac{1}{p}\right)+\frac{\lambda(p)}{p}\right)\right)\right) \\
&= \exp\left(y\left(-\frac{\gamma}{2}+M(3,4)\right)\right) \\
&< \exp\left(-y\cdot 0.2403\right) < \begin{cases} 1, & \text{if } 0 \leq y \leq 0.5005 \\ 0.8869, & \text{if } 0.4995 \leq y \leq 1.001. \end{cases}
\end{aligned}$$

Applying the explicit bounds calculated above and the triangle inequality in equation (4.11), for $0 \le y \le 0.5005$ we obtain

$$|f(y)| \le \frac{0.3148^2}{0.9063} + \frac{1.9782}{4 \cdot 0.9063^2} + \frac{0.1486}{0.9063} + \frac{0.5773 \cdot 0.3148}{0.9063^2} + \frac{0.5773^2}{2 \cdot 0.9063^3} < 1.3206,$$

and similarly for $0.4995 \le y \le 1.001$ we have

$$|f(y)| \le \frac{0.3892^2}{0.8856} + \frac{1.1327}{4 \cdot 0.8856^2} + \frac{0.1486}{0.8856} + \frac{0.2065 \cdot 0.3892}{0.8856^2} + \frac{0.2065^2}{2 \cdot 0.8856^3} < 0.8331.$$

Together with (4.10) this implies, for sufficiently large x:

$$1 + \frac{C(3,4)}{2} + \tilde{h}\left(\frac{2(k-3)}{\log\log x}\right) > 1 + \frac{0.0964}{2} - \frac{1}{2}(1.3206 + \frac{1}{2}(0.8869 \cdot 0.8331)) = 0.2031,$$

which leads to an admissible choice of $c = 0.2031$ in (4.9). $\square$

Bibliography

[1] A. Aigner, *Brüche als Summe von Stammbrüchen*, J. Reine Angew. Math. **214/215** (1964), 174–179.

[2] R. Arce-Nazario, F. Castro, and R. Figueroa, *On the number of solutions of $\sum_{i=1}^{11} \frac{1}{x_i} = 1$ in distinct odd natural numbers*, J. Number Theory **133** (2013), no. 6, 2036–2046.

[3] S. Baier, *A note on P-sets*, Integers **4** (2004), A13, 6 pages.

[4] C. Ballot and F. Luca, *On the sumset of the primes and a linear recurrence*, Acta Arith. **161** (2013), no. 1, 33–46.

[5] W. D. Banks and G. Martin, *Optimal primitive sets with restricted primes*, Integers **13** (2013), A69, 10 pages.

[6] A. T. Benjamin, B. Chen, and K. Kindred, *Sums of evenly spaced binomial coefficients*, Math. Mag. **83** (2010), no. 5, 370–373.

[7] A. S. Besicovitch, *On the density of certain sequences of integers*, Math. Ann. **110** (1935), no. 1, 336–341.

[8] L. Brenton and R. Hill, *On the Diophantine equation $1 = \sum 1/n_i + 1/\prod n_i$ and a class of homologically trivial complex surface singularities*, Pacific J. Math. **133** (1988), no. 1, 41–67.

[9] T. D. Browning and C. Elsholtz, *The number of representations of rationals as a sum of unit fractions*, Illinois J. Math. **55** (2011), no. 2, 685–696.

[10] E. R. Canfield, P. Erdős, and C. Pomerance, *On a problem of Oppenheim concerning "factorisatio numerorum"*, J. Number Theory **17** (1983), no. 1, 1–28.

[11] M.-C. Chang, *Short character sums for composite moduli*, J. Anal. Math. **123** (2014), 1–33.

[12] Y.-G. Chen, C. Elsholtz, and L.-L. Jiang, *Egyptian fractions with restrictions*, Acta Arith. **154** (2012), no. 2, 109–123.

[13] Y.-G. Chen and X.-G. Sun, *On Romanoff's constant*, J. Number Theory **106** (2004), no. 2, 275–284.

[14] J. G. van der Corput, *Over het vermoeden van de Polignac*, Simon Stevin **27** (1950), 99–105.

[15] R. Dedekind, *Über Zerlegungen von Zahlen durch ihre größten gemeinsamen Teiler*, in Gesammelte mathematische Werke, Zweiter Band (R. Fricke, E. Noether and Ö. Ore, eds.), Friedr. Vieweg & Sohn Akt.-Ges., Braunschweig, 1931, pp. 103–147. Available online at `http://resolver.sub.uni-goettingen.de/purl?PPN235693928`.

[16] A. Dubickas, *Sums of primes and quadratic linear recurrence sequences*, Acta Math. Sin. (Engl. Ser.) **29** (2013), no. 12, 2251–2260.

[17] M. Dunton and R. E. Grimm, *Fibonacci on Egyptian fractions*, Fibonacci Quart. **4** (1966), 339–354.

[18] C. Elsholtz, *Sums of k Unit Fractions*, Shaker Verlag, Aachen, 1999. Ph.D. Thesis, Technische Universität Darmstadt (1998), 109 pages.

[19] ______, *Sums of k unit fractions*, Trans. Amer. Math. Soc. **353** (2001), no. 8, 3209–3227.

[20] ______, *Egyptian fractions with odd denominators*, Q. J. Math. **67** (2016), no. 3, 425–430.

[21] C. Elsholtz, F. Luca, and S. Planitzer, *Romanov type problems*, Ramanujan J. (2018). Available online at `https://doi.org/10.1007/s11139-017-9972-8`.

[22] C. Elsholtz and S. Planitzer, *On Erdős and Sárközy's sequences with Property P*, Monatsh. Math. **182** (2017), no. 3, 565–575.

[23] ______, *The number of solutions of the Erdős-Straus Equation and sums of k unit fractions*, ArXiv e-prints, available online at `https://arxiv.org/abs/1805.02945` (May 2018).

[24] C. Elsholtz and J.-C. Schlage-Puchta, *On Romanov's constant*, Math. Z. **288** (2018), no. 3–4, 713–724.

[25] C. Elsholtz and T. Tao, *Counting the number of solutions to the Erdős-Straus equation on unit fractions*, J. Aust. Math. Soc. **94** (2013), no. 1, 50–105.

[26] P. Erdős, *Note on sequences of integers no one of which is divisible by any other*, J. London Math. Soc. **10** (1935), no. 2, 126–128.

[27] ______, *On sequences of integers no one of which divides the product of two others and on some related problems*, Mitt. Forsch.-Inst. Math. Mech. Univ. Tomsk **2** (1938), 74–82.

[28] ______, *Az $\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}=\frac{a}{b}$ egyenlet egész számú megoldásairól (On a Diophantine equation, in Hungarian)*, Mat. Lapok **1** (1950), 192–210.

[29] ______, *On integers of the form 2^k+p and some related problems*, Summa Brasil. Math. **2** (1950), 113–123.

[30] ______, *Some old and new problems on additive and combinatorial number theory*, in Combinatorial Mathematics: Proceedings of the Third International Conference (New York, 1985), *Ann. New York Acad. Sci.*, vol. 555, New York Acad. Sci., New York, 1989, pp. 181–186.

[31] ______, *Some of my favourite unsolved problems*, Math. Japon. **46** (1997), no. 3, 527–538.

[32] ______, *Some of my new and almost new problems and results in combinatorial number theory*, in Number theory (Eger, 1996; K. Győry, A. Pethő and V. T. Sós, eds.), de Gruyter, Berlin, 1998, pp. 169–180.

[33] P. Erdős and M. R. Murty, *On the order of $a \pmod p$*, in Number theory (Ottawa, ON, 1996; R. Gupta and K. S. Williams, eds.), *CRM Proc. Lecture Notes*, 19, Amer. Math. Soc., Providence, RI, 1999, pp. 87–97.

[34] P. Erdős and A. Sárközi, *On the divisibility properties of sequences of integers*, Proc. London Math. Soc. (3) **21** (1970), 97–101.

[35] P. Erdős and P. Turán, *Ein zahlentheoretischer Satz*, Mitt. Forsch.-Inst. Math. Mech. Univ. Tomsk **1** (1935), 101–103.

[36] ______, *Über die Vereinfachung eines Landauschen Satzes*, Mitt. Forsch.-Inst. Math. Mech. Univ. Tomsk **1** (1935), 144–147.

[37] L. Euler, *Letter to Goldbach*, Berlin, 16th Dec. 1752. Available online at `http://eulerarchive.maa.org/correspondence/letters/OO0879.pdf`.

[38] S. W. Golomb, *On certain nonlinear recurring sequences*, Amer. Math. Monthly **70** (1963), no. 4, 403–405.

[39] R. L. Graham, *On finite sums of reciprocals of distinct nth powers*, Pacific J. Math. **14** (1964), 85–92.

[40] ———, *On finite sums of unit fractions*, Proc. London Math. Soc. (3) **14** (1964), 193–207.

[41] ———, *Paul Erdős and Egyptian Fractions*, in Erdős Centennial, Bolyai Soc. Math. Stud., 25, (L. Lovász, I. Z. Ruzsa and V. T. Sós, eds.), János Bolyai Math. Soc., Budapest, 2013, pp. 289–309.

[42] R. K. Guy, *Unsolved problems in number theory*, third edn., Problem Books in Mathematics, Springer-Verlag, New York, 2010. Softcover reprint of hardcover third edn. 2004.

[43] L. Habsieger and X.-F. Roblot, *On integers of the form $p+2^k$*, Acta Arith. **122** (2006), no. 1, 45–50.

[44] L. Habsieger and J. Sivak-Fischler, *An effective version of the Bombieri-Vinogradov theorem, and applications to Chen's theorem and to sums of primes and powers of two*, Arch. Math. (Basel) **95** (2010), no. 6, 557–566.

[45] H. Halberstam and K. F. Roth, *Sequences*, Vol. 1, Oxford University Press, Oxford, 1966.

[46] R. R. Hall, *On the distribution of divisors of integers in residue classes* $(\mathrm{mod}\, k)$, J. Number Theory **2** (1970), 168–188.

[47] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, sixth edn., Oxford University Press, Oxford, 2008.

[48] G. Harman, *On the number of Carmichael numbers up to x*, Bull. London Math. Soc. **37** (2005), no. 5, 641–650.

[49] ———, *Watt's mean value theorem and Carmichael numbers*, Int. J. Number Theory **4** (2008), no. 2, 241–248.

[50] C. G. J. Jacobi, *Canon arithmeticus sive tabulae quibus exhibentur pro singulis numeris primis vel primorum potestatibus infra 1000 numeri ad datos indices et indices ad datos numeros pertinentes*, Berolinum, 1839.

[51] S. V. Konyagin, *Double exponential lower bound for the number of representations of unity by Egyptian fractions*, Math. Notes **95** (2014), no. 1-2, 277–281. Translation of Mat. Zametki **95** (2014), no. 2, 312–316.

[52] E. Landau, *Sur quelques problèmes relatifs à la distribution des nombres premiers*, Bull. Soc. Math. France **28** (1900), 25–38.

[53] A. Languasco and A. Zaccagnini, *Computing the Mertens and Meissel-Mertens constants for sums over arithmetic progressions*, Experiment. Math. **19** (2010), no. 3, 279–284. With an appendix by Karl K. Norton, computational results available online at `http://www.math.unipd.it/~languasc/Mertenscomput/Mqa/Msumfinalresults.pdf`.

[54] V.-A. Lebesgue, *Extrait des exercises d'analyse numérique*, Nouvelles annales de mathématiques 1re série **8** (1849), 347–353.

[55] K. S. E. Lee, *On the sum of a prime and a Fibonacci number*, Int. J. Number Theory **6** (2010), no. 7, 1669–1676.

[56] H. W. Lenstra, Jr. and C. Pomerance, *A rigorous time bound for factoring integers*, J. Amer. Math. Soc. **5** (1992), no. 3, 483–516.

[57] G. Martin, *Dense egyptian fractions*, Trans. Amer. Math. Soc. **351** (1999), no. 9, 3641–3657.

[58] G. Martin and C. Pomerance, *Primitive sets with large counting functions*, Publ. Math. Debrecen **79** (2011), no. 3-4, 521–530.

[59] X. Meng, *Large bias for integers with prime factors in arithmetic progressions*, Mathematika **64** (2018), no. 1, 237–252.

[60] ———, *Large bias for integers with prime factors in arithmetic progressions*, ArXiv e-prints, available online at `https://arxiv.org/abs/1607.01882v1` (July 2016).

[61] V. H. Moll, *Numbers and Functions: From a Classical-Experimental Mathematician's Point of View*, Amer. Math. Soc., Providence, RI, 2012.

[62] L. J. Mordell, *Diophantine equations*, Pure and Applied Mathematics, Vol. 30, Academic Press, London and New York, 1969.

[63] M. B. Nathanson, *Additive Number Theory - The Classical Bases*, Graduate Texts in Mathematics, vol. 164, Springer-Verlag, New York, 1996.

[64] J. Pintz, *A note on Romanov's constant*, Acta Math. Hungar. **112** (2006), no. 1-2, 1–14.

[65] A. de Polignac, *Recherches nouvelles sur les nombres premiers*, Comptes rendus hebdomadaires des séances de l'Académie des sciences **29** (1849), 397 –401 and 738–739.

[66] N. P. Romanoff, *Über einige Sätze der additiven Zahlentheorie*, Math. Ann. **109** (1934), no. 1, 668–678.

[67] L. A. Rosati, *Sull'equazione diofantea* $4/n = 1/x_1 + 1/x_2 + 1/x_3$, Boll. Un. Mat. Ital. (3) **9** (1954), 59–63.

[68] J. B. Rosser and L. Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois J. Math. **6** (1962), 64–94.

[69] C. Sándor, *On the number of solutions of the Diophantine equation* $\sum_{i=1}^{n} \frac{1}{x_i} = 1$, Period. Math. Hungar. **47** (2003), no. 1-2, 215–219.

[70] T. Schoen, *On a problem of Erdős and Sárközy*, J. Combin. Theory Ser. A **94** (2001), no. 1, 191–195.

[71] B. M. Stewart, *Theory of numbers*, second edn., The Macmillan Company, New York, Collier-Macmillan Limited, London, 1964.

[72] J. J. Sylvester, *On a Point in the Theory of Vulgar Fractions*, Amer. J. Math. **3** (1880), no. 4, 332–335.

[73] G. Tenenbaum, *Introduction à la théorie analytique et probabiliste des nombres*, third edn., Éditions Belin, Paris, 2008.

[74] ———, *Introduction to analytic and probabilistic number theory*, third edn., Graduate Studies in Mathematics, vol. 163, American Mathematical Society, Providence, RI, 2015.

[75] T. Xylouris, *Über die Nullstellen der Dirichletschen L-Funktionen und die kleinste Primzahl in einer arithmetischen Progression*, Ph.D. Thesis, 2011. Universität Bonn.

[76] K. Zsigmondy, *Zur Theorie der Potenzreste*, Monatsh. Math. Phys. **3** (1892), no. 1, 265–284.