The Pennsylvania State University
The Graduate School

ANALYTIC METHODS FOR DIOPHANTINE PROBLEMS

A Dissertation in
Mathematics
by

Jing-Jing Huang

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2012

The dissertation of Jing-Jing Huang was reviewed and approved* by the following:

Robert C. Vaughan
Professor of Mathematics
Dissertation Advisor, Co-Chair of Committee

Wen-Ching W. Li
Distinguished Professor of Mathematics
Co-Chair of Committee

W. Dale Brownawell
Distinguished Professor of Mathematics

Donald Richards
Professor of Statistics

Svetlana Katok
Professor of Mathematics
Director of Graduate Studies

*Signatures are on file in the Graduate School.

Abstract

In this dissertation, we are mainly concerned with the Diophantine equation

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

and its number of positive integer solutions $R_k(n;a)$. We begin with the binary case $k = 2$. Now the distribution of the function $R_2(n;a)$ is well understood. More precisely, by averaging over n, the first moment and second moment behaviors of $R_2(n;a)$ have been established. For instance, one of our results is

$$\sum_{\substack{n \le N \\ (n,a)=1}} R_2(n;a) = NP_2(\log N;a) + O_a(N \log^5 N),$$

where $P_2(\cdot;a)$ is a quadratic function whose coefficients depend on a. Furthermore, we have shown that, after normalisation, $R_2(n;a)$ satisfies Gaussian distribution, which is an analog of the classical theorem of Erdős and Kac,

$$\lim_{N\to\infty} \frac{1}{N} \mathrm{card} \left\{ \mathrm{n} \le \mathrm{N} : \frac{\log \mathrm{R}_2(\mathrm{n};\mathrm{a}) - (\log 3) \log\log \mathrm{n}}{(\log 3)\sqrt{\log\log \mathrm{n}}} \le \mathrm{z} \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\mathrm{z}} \mathrm{e}^{-\frac{\mathrm{t}^2}{2}} \mathrm{dt}.$$

On the other hand, we change the point of view and study the set of "exceptional numbers" that do not possess binary representations. Let $E_a(N)$ denote the number of $n \le N$ such that $R_2(n;a) = 0$. It is established that when $a \ge 3$ we have

$$E_a(N) \sim C(a) \frac{N(\log\log N)^{2^{m-1}-1}}{(\log N)^{1-1/2^m}},$$

with m defined in Chapter 4.

The next project would be to study the ternary case $k = 3$. While the conjecture, by Erdős, Straus and Schinzel, that for fixed $a \geq 4$, we have $R_3(n; a) > 0$ when n is sufficiently large, is still wide open, it is rather interesting to know the asymptotics for the mean value $\sum_{n \leq N} R_3(n; a)$.

Table of Contents

List of Figures

List of Symbols

$\ll$	$f(x) \ll g(x)$ means that there exists a positive constant C such that $f(x) \leq Cg(x)$ for x large enough or otherwise specified. If written $\ll_a$, we mean the underlying constant C depends on another parameter a.
$\phi(n)$	the Euler totient function
$d(n)$	the divisor function which counts the number of divisors of n
$\zeta(s)$	the Riemann zeta-function
χ	the Dirichlet character associated with some modulus.
$L(s,\chi)$	the Dirichlet L-function associated with the character χ
$R_k(n;a)$	the number of solutions to the Diophantine equation (1.1)
$S_k(N;a)$	the average $\sum_{\substack{n\leq N \\ (n,a)=1}} R(n;a)$
$T_k(N;a)$	the average $\sum_{n\leq N} R(n;a)$
$U_k(N)$	the average $\sum_{a\leq N}\sum_{\substack{n\leq N \\ (n,a)=1}} R(n;a)$

Acknowledgments

I would like to take this opportunity to thank my advisor, Robert Vaughan, for his patient guidance and constant encouragement, without which the current dissertation would not come out. His deep insight into and great vision of mathematics have brought me a deeper understanding of the wonderful subject — analytic number theory. I am also indebted to Winnie Li, whose extremely pertinent advices and great passion for mathematics have shaped me overall as a researcher. Besides, I thank Yao Chen for providing the data included in Figure 1.1. Finally, I could not achieve all of this without the love and support of my family: Jingyao, who maintains a sweet home for us; my parents, who show firm support of my academic career as always.

Chapter 1

Introduction and overview

1.1 Introduction to Egyptian fractions

A fundamental theme in mathematics is to study integral solutions to Diophantine equations. In this thesis, we are primarily interested in the following equation

$$\frac{a}{n} = \frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_k} \tag{1.1}$$

which has received extensive attention in the past few decades. Since Egyptians considered such representations, sums of unit fractions are sometimes called Egyptian fractions. A central question in this area is that what pair of a and n will entail a solution of the equation (1.1) in positive integers? and moreover how many are they? A more sensible question is that for a fixed positive integer a what is the minimum k such that for all sufficiently large n the equation (1.1) is always soluble in positive integers, namely $\frac{a}{n}$ can be written as the sum of k unit fractions.

Now if $a = 1$, then there is nothing to talk about and the minimum k is trivially 1. If $a = 2$, then all odd n need two unit fractions and hence the minimum k should be 2. In the case that $a = 3$, though slightly nontrivial, it is not hard to see that all prime numbers congruent to 1 mod 3 need three unit fractions. Now what happens if $a \geq 4$? This turns out to be much more interesting and in order to answer this question the following famous conjecture has been proposed.

Conjecture 1.1 (Erdős-Straus-Schinzel)**.** *For any integer $a \geq 4$, there exists an*

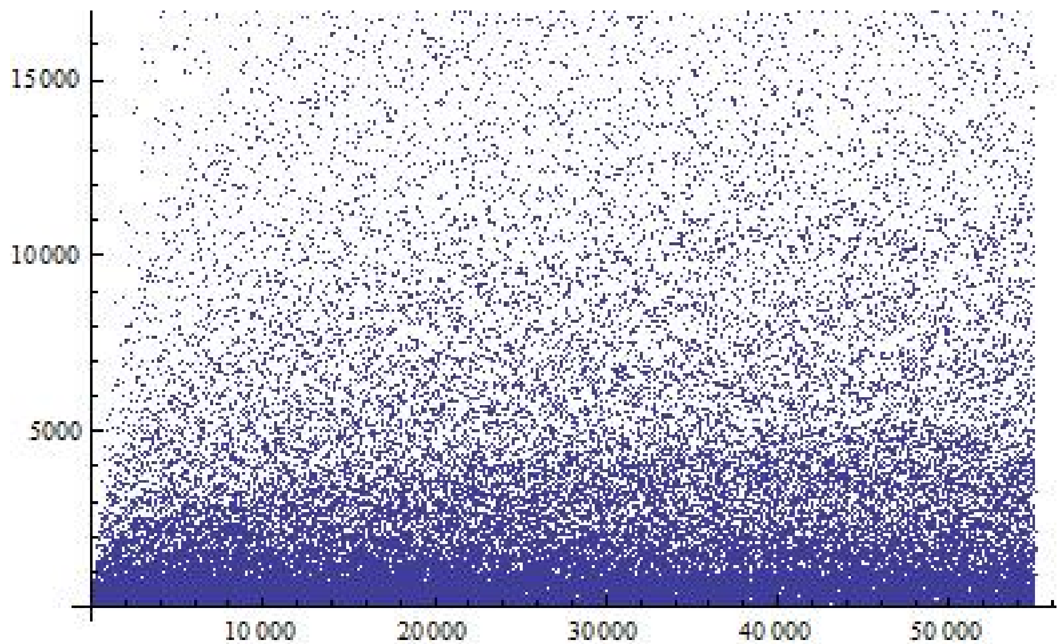

Figure 1.1. $R'_3(n;4)$ for $1 \leq n \leq 54999$

N_a, such that when $n > N_a$ the equation

$$\frac{a}{n} = \frac{1}{x} + \frac{1}{y} + \frac{1}{z} \tag{1.2}$$

is always solvable in positive integers.

The original conjecture for the case that $a = 4$ is due to Erdős and Straus in 1948 and the general case is due to Schinzel in 1956. Actually the versions presented by Erdős and Straus are slightly different. Straus conjectured that $\frac{4}{n}$ can be written as the sum of three distinct unit fractions when $n > 2$. As a matter of fact, these two versions are equivalent, the proof of which is included in Appendix A. Though trivial observation reveals that one can always find a desired representation when $a \leq k$, a surprising fact is that this still remains the best we can do in general! This means that we do not even know how to prove $\frac{a}{n}$ can be expressed as the sum of $a - 1$ unit fractions for large n.

Let $R_k(n;a)$ denote the number of positive integer solutions to the equation (1.1) and let $R'_k(n;a)$ be the corresponding quantity with the condition $x_1 \leq x_2 \leq \cdots \leq x_k$. In general the two functions should behave proportionally when a is fixed and n grows. The central difficulty of the conjecture lies in the fact that both $R_k(n;a)$ and $R'_k(n;a)$ oscillate a lot and do not admit any obvious pattern of growth *cf.* Figure 1.1. In particular, the figure suggests that $R'_3(n;4)$ grows like a logarithmic functions of n instead of a polynomial of n. This prevents any attempts to prove the conjecture by showing the asymptotic formula of $R_3(n;a)$ as n goes to infinity, which simply does not exist. Hence powerful tools like the Hardy-Littlewood circle method do not apply to this problem.

Remark 1.1. In 1974, William Webb [41] did some calculation to check how large those N_a in the Erdős-Straus-Schinzel conjecture should be. Especially, it is not clear at all how fast those optimal N_a grow as a tends to ∞. The original Erdős-Straus conjecture simply says that $N_4 = 1$. Webb's calculation suggests that $N_5 = N_6 = 1$, $N_7 = 2$, $N_8 = 241$. However, it is far from the truth that N_a is increasing. In fact it seems that N_a grows quite irregularly as a grows. For example, Webb's calculation also suggests that $N_{11} = 37$ and $N_{12} = 12241$.

Remark 1.2. The numerical verification for the original Erdős-Straus conjecture (i.e. when $a = 4$) has been carried out for $n \leq 10^{14}$ by Allen Swett in 1999 [34].

1.2 The number of counter examples

Put

$$\mathcal{E}_{a,k} = \{n \in \mathbb{N} : R_k(n;a) = 0\} \tag{1.3}$$

and the Erdős-Straus-Schinzel conjecture is equivalent to the assertion that $\mathcal{E}_{a,k}$ is a finite set for any $k \geq 3$ and $a \geq 1$. Let $E_{a,k}(N)$ be the number of elements in $\mathcal{E}_{a,k}$ that does not exceed N. Though the conjecture is out of reach in the current state of play, one may ask what can be said about the size of $E_{a,k}(N)$.

1.2.1 The binary case

In this subsection, we are primarily concerned with the number of counter examples $E_{a,2}(N)$ in the binary case $k = 2$. It obviously is true that both $\mathcal{E}_{1,2}$ and $\mathcal{E}_{2,2}$ are empty. However, things become much more interesting when $a \geq 3$.

In 1985, G. Hofmeister and P. Stoll [13] proved that the set $\mathcal{E}_{a,2}$ has asymptotic density 0, and more precisely that

$$E_{a,2}(N) \ll_a \frac{N}{(\log N)^{1/\phi(a)}}.$$

For $a = 5$ and $a \geq 7$ this bound is far from the truth. Their method is based on the observation that if the equation (4.1) is insoluble, then n is not divisible by any prime of the form $p \equiv -1 \pmod{a}$ *cf.* Lemma 4.3. Thus a simple application of Selberg's upper bound sieve gives the stated bound. However when $a = 5$ or $a \geq 7$ the bulk of the n deficient in such prime factors nevertheless have a representation.

We establish the precise asymptotic behavior of $E_{a,2}(N)$ in Theorem 4.1, which is the main result of [19].

For fixed $a \geq 3$, let $2^{\gamma_0} p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k}$ be the canonical decomposition of a and define m and δ by

$$2^m || \gcd(\delta, p_1 - 1, p_2 - 1, \cdots, p_k - 1)$$

and

$$\delta = \begin{cases} 0, & \text{if} \quad \gamma_0 \leq 1, \\ 2, & \text{if} \quad \gamma_0 \geq 2. \end{cases}$$

Then we have

$$E_{a,2}(N) \sim C(a) \frac{N(\log\log N)^{2^{m-1}-1}}{(\log N)^{1-1/2^m}},$$

where $C(a)$ is a positive constant depending only on a.

No lower bound for $E_a(N)$ has been drawn in [13] or in the literature, since most of the work in the literature is based on sieve methods, which are usually reluctant to produce nontrivial lower bounds. Also the bound in [13] is sharp only when $a = 3, 4, 6$ and gets worse as a tends to infinity. However, Theorem 4.1 shows that the worst bound is $\frac{N}{\sqrt{\log N}}$ and that for some a the bound is even as good as $\frac{N}{(\log N)^{1-\varepsilon}}$ for some fixed small $\varepsilon > 0$. More precisely, it is not hard to observe that for some random $a \geq 3$ the probability that $E_a(N)$ has order of magnitude $\frac{N(\log\log N)^{2^{m-1}-1}}{(\log N)^{1-1/2^m}}$ is $\frac{1}{2^m}$. Instead of sieve methods, we employ analytic methods based on Dirichlet series and, more importantly, it is necessary to elucidate the underlying group theoretic structure, which are the novel features of [19].

The proof of this result can be naturally split into two parts, i.e. the algebraic part and the analytic part.

The algebraic part is to understand the underlying structure of $\mathcal{E}_{a,2}$, and we embark on this in Section 4.2. This involves some interesting combinatorial and group theoretic arguments, which are the key components of the proof.

On the other hand, the main analytic input of the proof is fairly routine, and is based on an arithmetical application of a theorem of Delange (*cf.* Section 4.3). Delange's original theorem is stated as Lemma 4.10. In Lemma 4.11 we use standard methods on Dirichlet L-functions to transform Lemma 4.10 into a desired form for our application. Results of this kind at least date back to Hardy and Landau. Indeed, Hardy shows how to give the asymptotic formula for the number of integers that cannot be represented by the sum of two squares in his book [10]. The generating Dirichlet series essentially involves the square root of the Riemann zeta-function, and there are techniques going back to Landau for dealing with any fractional power of the zeta function, namely Dirichlet series with an algebraic pole. The novel feature of Lemma 4.11 is that it presents a unified version that deals with both algebraic and logarithmic poles.

On the whole, the key innovation in the proof of Theorem 4.1 is that we have a full understanding of the group theoretic structure of $\mathcal{E}_{a,2}$ and the combinatorial delicacy.

Remark 1.3. After we submitted our paper [19] to arXiv, Christian Elsholtz pointed out his Ph.D. thesis [5] and some subsequent unfinished work to us, in which he adumbrates some special cases of Theorem 4.1 but is not able to give a complete proof of the general case. Even in the case when a is a prime, his attempted proof is rather incomplete.

1.2.2 The ternary case

In order to understand the exceptional set $\mathcal{E}_{a,3}$ in the ternary case $k = 3$, it turns out that one needs to create a lot of solvable residue classes. This is shown in Appendix B. Combining these elementary results with sieve methods yields upper bounds on $\mathcal{E}_{a,3}$. The quality of the upper bound differs according to different sieve methods applied.

M. Nakayama [29] showed in 1940 that

$$E_{4,3}(N) \ll \frac{N}{\sqrt{\log N}}.$$

Later W. Webb [40] improved this to

$$E_{4,3}(N) \ll \frac{N}{(\log N)^{7/4}}.$$

The state of art result is due to R.C. Vaughan [38], who showed in 1970 that

$$E_a(N) \ll \frac{N}{\exp(c_a(\log N)^{\frac{2}{3}})},$$

where c_a depends at most on a.

The large improvement made by Vaughan is due to the application of the large sieve to Diophantine problems of this kind for the first time, whilst the previous two authors use only small sieves. Notice that the large sieve prevails since the number of solvable residue classes for each prime modulus, as is obtained in Appendix B, are unbounded.

Remark 1.4. By a finer adjustment to Vaughan's argument, Elsholtz worked out admissible values for c_a in his diploma thesis [4]. But, I found a serious mistake in his paper when I was doing my undergraduate thesis. Fortunately, we can fix that mistake and still obtain weaker admissible values for c_a as follows:

$$c_a = \frac{3}{4(2e^2a)^{\frac{1}{3}}}.$$

In particular, when $a = 4$, $c_4 = 0.1925$ is admissible.

The current state of play remains unsatisfactory in various aspects. The large sieve is presupposed to deal with the worst case, but our case might not be that bad; The large sieve only takes advantage of those residue classes with prime modulus, but we have more soluble residue classes with composite modulus; The large sieve only uses the prime modulus up to $\sqrt{N}$, while we have a bunch of primes greater than that.

It is not clear how to adjust Vaughan's method to prove $E_a(N) \ll N^{1-\epsilon}$ for some $\epsilon > 0$. The number of solvable residue classes obtained in Proposition B.8 is almost the best we can do, in the sense that it may subject to small improvement that is not good enough to improve the order of the magnitude of the bound but can just improve the constant c_a. It seems to me that as long as we insist on relying on the large sieve, it is very hard to improve Vaughan's result. On the other hand, if one seek to get new results along this line, one may want to devise a special sieve which is particularly suitable to our problem. Ideally that sieve should make full use of the information which we have, i.e. Proposition B.5, but which the large sieve disregards largely. Unfortunately, such a sieve simply does not exist within our current state of knowledge.

We have noted that those soluble residue classes will never cover square numbers (*cf.* Proposition B.7). Bounds like $E_a(N) \ll \sqrt{N}$ are the best that regular sieving process can produce. But the Erdős-Straus-Schinzel conjecture says that $E_a(N)$ is bounded. This reminds us that the ideal approach for this problem is to prove that all prime numbers are covered by the solvable residue classes produced in Proposition B.5. However, hardly anything is known on infinite covering congruent systems for integers, let alone such systems for the primes.

Our knowledge on the exceptional sets $E_{a,k}(N)$ for the cases $k > 3$ is similar to that of the case $k = 3$. Following Vaughan's argument, it is proved by C. Viola [39] that

$$E_{a,k}(N) \ll \frac{N}{\exp(c_a(\log N)^{1-\frac{1}{k-1}})},$$

for some constant c_a that depends on a. Z. Shan [32] improved this to

$$E_{a,k}(N) \ll \frac{N}{\exp\left(c_a(\log N)^{1-\frac{1}{k}}\right)}.$$

Finally, by looking into the parametric solutions closely and producing more solvable residue classes, C. Elsholtz was able to show

$$E_{a,k}(N) \ll \frac{N}{\exp\left(c_a(\log N)^{1-\frac{1}{2^{k-1}-1}}\right)}.$$

Notice that none of the above results improves Vaughan's original bound when

$k = 3$.

1.3 Averages and distributions

It is a convention in Mathematics that if one cannot understand well an individual object, one may simply consider a family of such objects and try to investigate the average behavior of certain properties and even the distribution of these properties among the family. Not surprisingly we will follow this convention.

1.3.1 The binary case

We have proved the following mean value theorem for $R_2(n;a)$, *cf.* Theorem 2.1, which is part of the results in [17].

$$\sum_{\substack{n\le N\\(n,a)=1}} R_2(n;a) = NP_2(\log N;a) + O_a(\sqrt{N}\log^5 N)$$

where $P_2(\cdot;a)$ is a quadratic function whose coefficients depend on a.

One major idea in the proof of Theorem 2.1 is to express $R_2(n;a)$ as coefficients of a linear combination of Dirichlet L-functions and therefore powerful analytic methods can be launched. Applying this idea, I have also proved a similar mean value result for the Diophantine equation $axy - x - y = n$ where a is fixed and n varies. See Chapter 5 for more details and this result is also published in [14].

On the other hand, by further averaging over a, we have shown by a combination of elementary and analytic methods that

$$\sum_{a}\sum_{\substack{n\le N\\(n,a)=1}} R_2(n;a) = \frac{1}{4}CN(\log N)^3 + O\big(N(\log N)^2\big),$$

where $C = \prod_p \big(1 - 3p^{-2} + 2p^{-3}\big)$.

This is included as Theorem 2.2 which improves upon an earlier result by Croot, Dobbs, Friedlander, Hetzel and Pappalardi [2] in which the weaker error

$$O(N(\log N)^3)/\log\log N)$$

is obtained.

Moreover, in Chapter 3 (*cf.* [18]), we are able to establish the second moment estimate for $R_2(n;a)$, from which we have deduced that after normalization, $R_2(n;a)$ admits a Gaussian distribution, which is an analog of the classical theorem of Erdős and Kac. More precisely, we show:

For fixed positive integer a, we have, for every $N \in \mathbb{N}$ with $N \geq 2$,

$$\sum_{\substack{n\leq N\\(n,a)=1}} \left| R_2(n;a) - \frac{1}{\phi(a)} \sum_{\substack{\chi \bmod a\\ \chi^2=\chi_0}} \bar{\chi}(-n) \sum_{u|n^2} \chi(u) \right|^2 \ll_a N \log^2 N,$$

where $\ll_a$ indicates that the implicit constant depends at most on a, and where χ_0 denotes the principal character modulo a.

From this we readily deduce that when a is fixed, the normal order of $\log R_2(n;a)$ as a function of n is $(\log 3)\log\log n$. With a bit more work, we establish the full distribution of $R_2(n;a)$.

For fixed positive integer a, we have

$$\lim_{N\to\infty} \frac{1}{N} \mathrm{card}\left\{ \mathrm{n} \leq \mathrm{N} : \frac{\log \mathrm{R}_2(\mathrm{n};\mathrm{a}) - (\log 3)\log\log \mathrm{n}}{(\log 3)\sqrt{\log\log \mathrm{n}}} \leq \mathrm{z} \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\mathrm{z}} \mathrm{e}^{-\frac{\mathrm{t}^2}{2}} \mathrm{dt}.$$

From here we see that $\log R_2(n;a)$ admits the normal distribution with mean $\log 3 \log\log n$ and standard deviation $\log 3\sqrt{\log\log n}$.

Lastly for completeness, we also obtain the regular second moment

$$\sum_{\substack{n\leq N\\(n,a)=1}} R(n;a)^2 = N P_8(\log N; a) + O_a(N^{35/54+\varepsilon})$$

where $P_8(\cdot\,; a)$ is a degree 8 polynomial with coefficients depending on a, and its leading coefficient is

$$\frac{1}{8!a^2} \prod_{p|a} \left(1-\frac{1}{p}\right)^7 \prod_{p\nmid a} \left(1+\frac{6}{p}+\frac{1}{p^2}\right)\left(1-\frac{1}{p}\right)^6.$$

The proof of this result utilizes the twelfth moment estimate of the Dirichlet L-functions on the critical line ([24] and [25]).

1.3.2 The ternary case

The ternary case, $k = 3$, is much harder and resistant to attacks. Recently Christian Elsholtz and Terrence Tao [7] showed among other things that

$$N(\log N)^2 \ll \sum_{p \le N} R_3(p; a) \ll_a N(\log N)^2 \log\log N,$$

where the summation runs over prime numbers. The double logarithmic factor should not be there, though they cannot remove it due to some technical difficulties in applying the Brun-Titchmarsh inequality.

I believe that a more interesting and more essential question is to estimate the general mean value of $R_3(n; a)$, and I have conjectured that

Conjecture 1.2.

$$\sum_{n \le N} R_3(n; a) = C_a N(\log N)^9 + o(N(\log N)^9)$$

Even the stronger statement is very likely to be true

Conjecture 1.3.

$$\sum_{n \le N} R_3(n; a) = N P_9(\log N; a) + O_a(N^{1-\delta}),$$

where $\delta > 0$ and $P_9(x; a)$ is a polynomial of degree 9 in x whose coefficients depend on a.

For the general case, I have also raised similar conjectures with an explicit conjectural exponent of $\log N$. The problems here are closely related with some corresponding problems about Manin's conjecture. More precisely, Roger Heath-Brown [12] studied the density of rational points on the projective variety

$$\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \frac{1}{x_4} = 0.$$

Motivated by his work, in an ongoing project [15] I am able to show the lower bound

$$\sum_{n\leq N} R_3(n;a) \gg_a N(\log N)^9$$

which is of the correct order of magnitude as conjectured above. The exponent 9 should be compared with the exponent 6 that appears in [12]. In spite of the similarities between these problems, the problem I am considering here has a different nature from the one in [12], in that Heath-Brown deals with the case that all the variables x_i are bounded by N, while my problem restricts only one variable. Heath-Brown reduces the lower bound of his problem to the counting of the number of lattice points in some bounded regular region, however, after working out the details for my problem, it turns out that I need to count the number of lattice points in some hyperbolic region with cusps and to express that quantity explicitly in terms of the parameters appearing in the problem in order for further averaging over them, and this brings up some substantial difficulties. This also explains the difference in the exponents of logarithmic powers in the two problems.

Unlike that in [12], the upper bound here seems genuinely harder and will be a future project.

Chapter 2

Mean Value Theorems for Binary Egyptian Fractions I

2.1 Introduction

The solubility of the Diophantine equation

$$\frac{a}{n} = \frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_k}, \tag{2.1}$$

in positive integers $x_1, x_2, \ldots, x_k$ has a long history. See, for example, Guy [8] for a detailed survey on this topic and a more extensive bibliography. When $k \geq 3$ it is still an open question as to whether the equation is always soluble provided that $n > n_0(a, k)$. When $k = 3$ the strongest result in this direction is Vaughan [37], [38] (see also Shan [32], Viola [39] and Elsholtz [6] for general k). In this chapter we are concerned with the case $k = 2$. In that case it is known that for any given $a > 2$ there are infinitely many n for which the equation is insoluble. For example, the criterion enunciated in the first paragraph of Section 2.3 shows that no n with all its prime factors p of the form $p \equiv 1 \pmod{a}$ has such a representation. However the number

$$R(n; a) = \operatorname{card}\left\{(x, y) \in \mathbb{N}^2 : \frac{a}{n} = \frac{1}{x} + \frac{1}{y}\right\} \tag{2.2}$$

of representations has an interesting and complicated multiplicative structure and can be studied in a number of ways. Here we consider various averages

$$S(N;a) = \sum_{\substack{n\le N\\(n,a)=1}} R(n;a), \tag{2.3}$$

$$T(N;a) = \sum_{n\le N} R(n;a)$$

and

$$U(N) = \sum_{a} S(N;a).$$

Croot et al. [2] have shown that

$$U(N) = \frac{1}{4}CN(\log N)^3 + O\left(\frac{N(\log N)^3}{\log\log N}\right),$$

and in Theorem 2.2 below we obtain a significant strengthening. However, in the main result of this chapter, Theorem 2.1, below, we show that it is possible to obtain a strong asymptotic formula without the necessity of averaging over a.

Theorem 2.1.

$$S(N;a) = \frac{3}{\pi^2 a}\left(\prod_{p|a}\frac{p-1}{p+1}\right)N\big((\log N)^2 + c_1(a)\log N + c_0(a)\big) + \Delta(N;a)$$

where

$$c_1(a) = 6\gamma - 4\frac{\zeta'(2)}{\zeta(2)} - 2 + \sum_{p|a}\frac{6p+2}{p^2-1}\log p$$

and

$$c_0(a) = -2(\log a)^2 - 4(\log a)\sum_{p|a}\frac{\log p}{p-1} + O(a\phi(a)^{-1}\log a),$$

and

$$\Delta(N;a) \ll N^{\frac{1}{2}}(\log(N))^5\frac{a}{\phi(a)}\prod_{p|a}\left(1-p^{-1/2}\right)^{-1}$$

uniformly for $N\ge 4$ and $a\in\mathbb{N}$.

Since

$$T(N;a) = \sum_{d|a} S\left(\frac{N}{d};\frac{a}{d}\right)$$

it is a straightforward exercise to obtain the corresponding asymptotic expansion for T.

The main novelty is the employment, for the first time in this area, of complex analytic techniques from multiplicative number theory. One may speculate on the utility of assuming the Generalised Riemann Hypothesis (GRH) in possibly improving the error term here significantly. This is unlikely with the proof in its present form, since the the main theoretical input from Dirichlet L–functions is *via* Lemma 2.7 below and the bounds there are at least as strong as can be established on GRH apart possibly from the power of the logarithm. However, in view of the aforementioned criterion in Section 2.2, the underlying problem has some affinity with the generalised divisor problem in the case of $d_3(n)$ and it is conceivable that, by pursuing methods related to that problem, an error bound of the form

$$O(N^\theta)$$

can be obtained with

$$\frac{1}{3} < \theta < \frac{1}{2}.$$

Theorem 2.2. *We have*

$$U(N) = \frac{1}{4}CN(\log N)^3 + O\big(N(\log N)^2\big),$$

where $C = \prod_p \left(1 - 3p^{-2} + 2p^{-3}\right)$.

Remark 2.1. In the Zentralblatt review of [2], the reviewer adumbrates a proof of a result somewhat weaker than Theorem 2.2.

This chapter is organized as follows. In Section 2.2, we state several lemmas which are needed in the proof of Theorem 2.1. In Section 2.3, we present an analytic proof of Theorem 2.1 based on Dirichlet L-functions. And in Section 2.4, an essentially elementary proof of Theorem 2.2 is given. Finally, in Section 2.5, we list some open questions in this area.

2.2 Preliminary Lemmas

We state several lemmas before embarking on the proof of Theorem 2.1. The content of Lemma 2.3 can be found, for example, in Corollary 1.17 and Theorem 6.7 of Montgomery and Vaughan [28], and Lemma 2.4 can be deduced from Theorem 4.15 of Titchmarsh [36] with $x = y = (|t|/2\pi)^{1/2}$.

Lemma 2.3. *When $\sigma \geq 1$ and $|t| \geq 2$, we have*

$$\frac{1}{\log |t|} \ll \zeta(\sigma + it) \ll \log |t|.$$

Lemma 2.4. *When $0 \leq \sigma \leq 1$ and $|t| \geq 2$, we have*

$$\zeta(\sigma + it) \ll |t|^{\frac{1-\sigma}{2}} \log(|t|).$$

Lemma 2.5. *Let χ be a non-principle character modulo a and $s = \sigma + it$ and assume that $t \in \mathbb{R}$. Then*

$$L(s, \chi) \ll \log(a(2 + |t|)), \text{ when } \sigma \geq 1$$

and

$$L(s, \chi) \ll (a|t|)^{\frac{1-\sigma}{2}+\varepsilon}, \text{ when } \frac{1}{2} \leq \sigma \leq 1.$$

Proof. The first part follows from Lemma 10.15 of Montgomery and Vaughan [28]. Now suppose that χ is primitive. Then by Corollary 10.10 of Montgomery and Vaughan [28],

$$L(s, \chi) \ll (a|t|)^{\frac{1}{2}-\sigma} \log(a(2 + |t|))$$

when $\sigma \leq 0$. Then by the convexity principle for Dirichlet series, for example as described in Titchmarsh [36] (cf. Exercise 10.1.19 of Montgomery and Vaughan [28]),

$$L(s, \chi) \ll (a|t|)^{\frac{1-\sigma}{2}+\varepsilon}$$

when $0 \leq \sigma \leq 1$. The proof is completed by observing that if $\frac{1}{2} \leq \sigma \leq 1$ and χ

modulo a is induced by the primitive character χ^* with conductor q, then

$$L(s,\chi) = L(s,\chi^*)\prod_{\substack{p|a \\ p\nmid q}}(1-\chi^*(p)p^{-s}) \ll |L(s,\chi^*)|2^{\omega(a)}.$$

□

Lemma 2.6. *Let $T \geq 2$, then we have*

$$\int_{-T}^{T} |\zeta(\tfrac{1}{2}+it)|^4 dt \sim \tfrac{1}{\pi^2} T \log^4 T$$

and

$$\sideset{}{^*}\sum_{\substack{\chi \\ \bmod a}} \int_{-T}^{T} |L(\tfrac{1}{2}+it,\chi)|^4 dt \ll \phi(a) T(\log(aT))^4,$$

where $\sum^*$ *indicates that the sum is over the primitive characters modulo a.*

The first formula here is due to Ingham [20] and the second is Theorem 10.1 of Montgomery [27].

Lemma 2.7. *Let $T \geq 2$, then*

$$\sum_{\substack{\chi \\ \bmod a}} \int_{-T}^{T} |L(\tfrac{1}{2}+it,\chi)|^4 dt \ll aT(\log(aT))^4,$$

Proof. Suppose that the character χ modulo a is induced by the primitive character χ^* with conductor q. Then the L–function in the integrand in modulus is

$$|L(\tfrac{1}{2}+it,\chi^*)\textstyle\prod_{p|a,p\nmid q}(1-\chi^*(p)p^{-1/2-it})| \leq |L(\tfrac{1}{2}+it,\chi^*)|\prod_{p|a/q}(1+p^{-1/2}).$$

Hence by the previous lemma

$$\sum_{\substack{\chi \\ \bmod a}} \int_{-T}^{T} |L(\tfrac{1}{2}+it,\chi)|^4 dt \ll T(\log(aT))^4 \textstyle\sum_{q|a}\phi(q)\prod_{p|a/q}(1+p^{-1/2})^4.$$

The sum here is

$$\sum_{q|a} \phi(q) \prod_{p|a/q} \left(1 + p^{-\frac{1}{2}}\right)^4$$
$$= \prod_{p^k \| a} \left(\left(1 + p^{-\frac{1}{2}}\right)^4 + \sum_{h=1}^{k-1} \phi(p^h) \left(1 + p^{-\frac{1}{2}}\right)^4 + \phi(p^k) \right)$$
$$= a \prod_{p|a} \left(1 + p^{-1}\left((1 + p^{-1/2})^4 - 1\right)\right) \ll a.$$

□

2.3 Proof of Theorem 2.1

Without loss of generality, we can assume $a \leq 2N$, since $R(n;a) = 0$ whenever $a > 2n$. Now we rewrite the equation $\frac{a}{n} = \frac{1}{x} + \frac{1}{y}$ in the form

$$(ax - n)(ay - n) = n^2.$$

After the change of variables $u = ax - n$ and $v = ay - n$, it follows that $R(n;a)$ is the number of ordered pairs of natural numbers u, v such that $uv = n^2$ and $u \equiv v \equiv -n \pmod{a}$.

Under the assumption that $(n, a) = 1$, $R(n;a)$ can be further reduced to counting the number of divisors u of n^2 with $u \equiv -n \pmod{a}$. Now the residue class $u \equiv -n \pmod{a}$ is readily isolated *via* the orthogonality of the Dirichlet characters χ modulo a. Thus we have

$$S(N;a) = \sum_{\substack{n \leq N \\ (n,a)=1}} R(n;a)$$
$$= \sum_{\substack{n \leq N \\ (n,a)=1}} \frac{1}{\phi(a)} \sum_{\substack{\chi \\ \text{mod } a}} \bar{\chi}(-n) \sum_{u|n^2} \chi(u)$$
$$= \frac{1}{\phi(a)} \sum_{\substack{\chi \\ \text{mod } a}} \bar{\chi}(-1) \sum_{n \leq N} \bar{\chi}(n) \sum_{u|n^2} \chi(u),$$

where the condition $(n, a) = 1$ is taken care of by the character $\bar{\chi}(n)$.

Let

$$a_n(\chi) = \bar{\chi}(n) \sum_{u|n^2} \chi(u). \tag{2.4}$$

Then we have

$$S(N;a) = \frac{1}{\phi(a)} \sum_{\substack{\chi \\ \bmod a}} \bar{\chi}(-1) \sum_{n \le N} a_n(\chi).$$

We analyze this expression through the properties of the Dirichlet series

$$f_\chi(s) = \sum_{n=1}^{\infty} \frac{a_n(\chi)}{n^s}.$$

The condition $u|n^2$ can be rewritten uniquely as $u = n_1 n_2^2$ and $n = n_1 n_2 n_3$ with n_1 square-free. Hence, for $\sigma > 1$ we have

$$\begin{aligned} f_\chi(s) &= \sum_{n=1}^{\infty} \frac{\bar{\chi}(n)}{n^s} \sum_{u|n^2} \chi(u) \\ &= \sum_{n_1,n_2,n_3=1}^{\infty} \mu(n_1)^2 \frac{\bar{\chi}(n_1 n_2 n_3)\chi(n_1 n_2^2)}{n_1^s n_2^s n_3^s} \\ &= \sum_{n_1=1}^{\infty} \frac{\mu(n_1)^2 \chi_0(n_1)}{n_1^s} \sum_{n_2=1}^{\infty} \frac{\chi(n_2)}{n_2^s} \sum_{n_3=1}^{\infty} \frac{\bar{\chi}(n_3)}{n_3^s} \end{aligned}$$

and so

$$f_\chi(s) = \frac{L(s,\chi_0)}{L(2s,\chi_0)} L(s,\chi) L(s,\bar{\chi}), \tag{2.5}$$

where χ_0 is the principal character modulo a, and this affords an analytic continuation of f_χ to the whole of $\mathbb{C}$.

By a quantitative version of Perron's formula, as in Theorem 5.2 of Montgomery and Vaughan [28] for example, we obtain

$$\sideset{}{'}\sum_{n \le N} a_n(\chi) = \frac{1}{2\pi i} \int_{\sigma_0 - iT}^{\sigma_0 + iT} f_\chi(s) \frac{N^s}{s} ds + R(\chi),$$

where $\sigma_0 > 1$ and

$$R(\chi) \ll \sum_{\substack{\frac{N}{2}<n<2N \\ n\neq N}} |a_n(\chi)| \min\left(1, \frac{N}{T|n-N|}\right) + \frac{4^{\sigma_0}+N^{\sigma_0}}{T}\sum_{n=1}^{\infty}\frac{|a_n(\chi)|}{n^{\sigma_0}}.$$

Here $\sum'$ means that when N is an integer, the term $a_N(\chi)$ is counted with weight $\frac{1}{2}$.

Let $\sigma_0 = 1 + \frac{1}{\log N}$. By (2.4) we have $|a_n(\chi)| \leq d(n^2)$. Thus

$$\sum_{n=1}^{\infty}\frac{|a_n(\chi)|}{n^{\sigma_0}} \ll \zeta(\sigma_0)^3 \ll (\log N)^3$$

and so $R(\chi) \ll_\varepsilon N^{1+\varepsilon}T^{-1}$, for any $\varepsilon > 0$. Hence

$$\sum_{n\leq N} a_n(\chi) = \frac{1}{2\pi i}\int_{\sigma_0-iT}^{\sigma_0+iT} f_\chi(s)\frac{N^s}{s}ds + O\left(\left(\frac{N}{T}+1\right)N^{\epsilon}\right).$$

The error term here is

$$\ll N^{\varepsilon}$$

provided that

$$T \geq N.$$

The integrand is a meromorphic function in the complex plane and is analytic for all s with $\Re s \geq \frac{1}{2}$ except for a pole of finite order at $s=1$. Suppose that $T \geq 4$. By the residue theorem

$$\frac{1}{2\pi i}\int_{\sigma_0-iT}^{\sigma_0+iT} f_\chi(s)\frac{N^s}{s}ds = \mathrm{Res}_{s=1}\left(f_\chi(s)\frac{N^s}{s}\right) +$$
$$\frac{1}{2\pi i}\left(\int_{\sigma_0-iT}^{\frac{1}{2}-iT} + \int_{\frac{1}{2}-iT}^{\frac{1}{2}+iT} + \int_{\frac{1}{2}+iT}^{\sigma_0+iT}\right)\frac{L(s,\chi_0)L(s,\chi)L(s,\bar{\chi})N^s}{L(2s,\chi_0)s}ds$$

We have $L(s,\chi_0) = \zeta(s)\prod_{p|a}(1-p^{-s})$. Hence, by Lemmas 1, 2 and 3 and the fact that $\prod_{p|a}(1-p^{-s}) \ll \log\log a$ when $\sigma \geq 1$, the contribution from the horizontal paths is

$$\ll (\log aT)^2(\log T)(\log\log a))NT^{-1} + T^{-1}(aT)^{\varepsilon}\int_{1/2}^{1}(aT)^{\frac{3(1-\sigma)}{2}}N^{\sigma}d\sigma$$
$$\ll T^{-1}(aT)^{\varepsilon}N + T^{-1}(aT)^{3/4+\varepsilon}N^{1/2}$$

and provided that $a \leq 2N$ and $T \geq N^{10}$ this is

$$\ll N^{-1}.$$

On the other hand, by Lemma 1 the contribution from the vertical path on the right is bounded by

$$N^{\frac{1}{2}}\left(\prod_{p|a}(1-p^{-\frac{1}{2}})^{-1}\right)(\log T)\sum_{2^k\leq T}2^{-k}I(k,\chi)$$

where

$$I(k,\chi) = \int_{-2^{k+1}}^{2^{k+1}}|\zeta(\tfrac{1}{2}+it)L(\tfrac{1}{2}+it,\chi)L(\tfrac{1}{2}+it,\bar{\chi})|dt.$$

By Lemmas 4 and 5 and Hölder's inequality

$$\sum_{\substack{\chi \\ \bmod a}}\frac{1}{2\pi i}\int_{\frac{1}{2}-iT}^{\frac{1}{2}+iT}\frac{L(s,\chi_0)L(s,\chi)L(s,\bar{\chi})N^s}{L(2s,\chi_0)s}ds$$
$$\ll N^{\frac{1}{2}}\left(\prod_{p|a}(1-p^{-\frac{1}{2}})^{-1}\right)(\log T)\sum_{2^k\leq T}a(k+\log a)^3$$
$$\ll N^{\frac{1}{2}}\left(\prod_{p|a}(1-p^{-\frac{1}{2}})^{-1}\right)a(\log N)^5$$

on taking

$$T = N^{10}.$$

Thus we have shown that

$$S(N;a) = \frac{1}{\phi(a)}\sum_{\substack{\chi \\ \bmod a}}\bar{\chi}(-1)\mathrm{Res}_{s=1}\left(f_{\chi}(s)\frac{N^s}{s}\right) + \Delta(N;a)$$

where

$$\Delta(N;a) \ll N^{\frac{1}{2}}(\log N)^5 \frac{a}{\phi(a)} \prod_{p|a} \left(1 - p^{-1/2}\right)^{-1}$$

It remains to compute the residue at $s = 1$.

By (2.5) there are naturally two cases, namely, $\chi \neq \chi_0$ and $\chi = \chi_0$. When $\chi \neq \chi_0$ the integrand has a simple pole at $s = 1$ and the residue is

$$\mathrm{Res}_{s=1} \left(\frac{L(s,\chi_0)L(s,\chi)L(s,\bar{\chi})N^s}{L(2s,\chi_0)s} \right) = \frac{6N}{\pi^2} \left(\prod_{p|a} \frac{p}{p+1} \right) |L(1,\chi)|^2.$$

It is useful to have some understanding of the behavior of

$$\frac{1}{\phi(a)} \sum_{\substack{\chi \neq \chi_0 \\ \bmod a}} \bar{\chi}(-1)|L(1,\chi)|^2.$$

Let $x = a^3$. Then for non-principal characters χ modulo a, by Abel summation

$$L(1,\chi) = \sum_{n \leq x} \frac{\chi(n)}{n} + O(a^{-2}).$$

Hence

$$\frac{1}{\phi(a)} \sum_{\substack{\chi \neq \chi_0 \\ \bmod a}} \bar{\chi}(-1)|L(1,\chi)|^2$$

$$= \frac{1}{\phi(a)} \sum_{\substack{\chi \neq \chi_0 \\ \bmod a}} \bar{\chi}(-1) \left| \sum_{n \leq x} \frac{\chi(n)}{n} \right|^2 + O\left(a^{-1}\right).$$

The main term on the right is

$$\frac{1}{\phi(a)} \sum_{\substack{\chi \\ \bmod a}} \bar{\chi}(-1) \left| \sum_{n \leq x} \frac{\chi(n)}{n} \right|^2 - \frac{1}{\phi(a)} \left(\sum_{\substack{n \leq x \\ (n,a)=1}} \frac{1}{n} \right)^2.$$

We have

$$\begin{aligned}\sum_{\substack{n\le x\\(n,a)=1}}\frac{1}{n}&=\sum_{m|a}\frac{\mu(m)}{m}\sum_{n\le x/m}\frac{1}{n}\\&=\sum_{m|a}\frac{\mu(m)}{m}(\log(x/m)+\gamma+O(m/x))\\&=\frac{\phi(a)}{a}\left(\log x+\gamma+\sum_{p|a}\frac{\log p}{p-1}\right)+O(d(a)/x).\end{aligned}$$

Hence the second term above is

$$-\frac{\phi(a)}{a^2}\left(\log x+\gamma+\sum_{p|a}\frac{\log p}{p-1}\right)^2+O(1/a).$$

The first term above is

$$\sum_{\substack{m,n\le x\\(mn,a)=1\\a|m+n}}\frac{1}{mn}.$$

The terms with $m=n$ contribute

$$\sum_{\substack{m\le x\\(m,a)=1\\a|2m}}\frac{1}{m^2}\ll a^{-2}$$

and this can be collected in the error term. The remaining terms are collected together so that $m+n=ak$, $m\ne n$ and $k\le\frac{2x}{a}$. If necessary by interchanging m and n we can suppose that $m<n$. Thus the above is

$$\sum_{1\le k\le 2x/a}\sum_{\substack{m\le x\\0<ak-m\le x\\m\le ak/2\\(m,a)=1}}\frac{2}{m(ak-m)}.$$

On interchanging the order of summation this becomes

$$\sum_{\substack{m\le x\\(m,a)=1}} \frac{2}{m} \sum_{2m/a<k\le (x+m)/a} \frac{1}{ak-m}.$$

We now divide the sum over m according as $m > a/2$ or $m \le a/2$. In the former case the inner sum can be written as the Stieltjes integral

$$\int_{(2m/a)+}^{(x+m)/a+} \frac{d\lfloor\alpha\rfloor}{a\alpha - m} = \frac{\lfloor (x+m)/a\rfloor}{x} - \frac{\lfloor 2m/a\rfloor}{m} + \int_{2m/a}^{(x+m)/a} \frac{a\lfloor\alpha\rfloor}{(a\alpha-m)^2}d\alpha.$$

Since $m \le x$ the first term is $\ll 1/a$, and the second term is 0 unless $m \ge \frac{a}{2}$, in which case it is $\ll 1/a$. Thus these terms contribute $\ll (\log a)/a$ in total. The integral here is

$$\int_{2m/a}^{(x+m)/a} \frac{a\alpha - m - a(\alpha - \lfloor\alpha\rfloor) + m}{(a\alpha-m)^2}d\alpha = a^{-1}\log(x/m) + O(1/a).$$

Thus the contribution to our sum is

$$a^{-1} \sum_{\substack{a/2<m\le x\\(m,a)=1}} \frac{2}{m}\log(x/m) + O\big((\log a)a^{-1}\big).$$

When $m \le a/2$ the sum over k becomes instead

$$\int_{1-}^{(x+m)/a+} \frac{d\lfloor\alpha\rfloor}{a\alpha - m} = \frac{\lfloor (x+m)/a\rfloor}{x} + \int_{1}^{(x+m)/a} \frac{a\lfloor\alpha\rfloor}{(a\alpha-m)^2}d\alpha.$$

The first term is $\ll 1/a$ and the integral is

$$\int_{1}^{(x+m)/a} \frac{a\alpha - m - a(\alpha - \lfloor\alpha\rfloor) + m}{(a\alpha-m)^2}d\alpha = a^{-1}\log(x/(a-m)) + O(1/a).$$

Thus we have shown that

$$\frac{1}{\phi(a)} \sum_{\substack{\chi\ne\chi_0\\ \bmod a}} \bar\chi(-1)|L(1,\chi)|^2 = a^{-1} \sum_{\substack{m\le x\\(m,a)=1}} \tfrac{2}{m}\log\tfrac{x}{m} - a^{-1} \sum_{\substack{m\le a/2\\(m,a)=1}} \tfrac{2}{m}\log\tfrac{a-m}{m}$$

$$-\frac{\phi(a)}{a^2}\left(\log x+\gamma+\sum_{p|a}\frac{\log p}{p-1}\right)^2+O((\log a)a^{-1}).$$

The first sum on the right is

$$2a^{-1}\sum_{k|a}\frac{\mu(k)}{k}\sum_{n\le x/k}n^{-1}\log(x/kn)$$

and this is readily seen to be

$$a^{-1}\sum_{k|a}\frac{\mu(k)}{k}\Big((\log(x/k))^2+2\gamma\log(x/k)+C\Big)+O(d(a)/(ax))$$

for a suitable constant C. Here the main term is

$$\frac{\phi(a)}{a^2}\left(\Big(\log x+\gamma+\sum_{p|a}\frac{\log p}{p-1}\Big)^2+O\big((\log\log(3a))^2\big)\right).$$

Hence, we have

$$\frac{1}{\phi(a)}\sum_{\substack{\chi\ne\chi_0\\ \bmod a}}\bar\chi(-1)|L(1,\chi)|^2=-a^{-1}\sum_{\substack{m\le a/2\\ (m,a)=1}}\tfrac{2}{m}\log\tfrac{a-m}{m}+O(a^{-1}\log 2a).$$

The sum over m is

$$\begin{aligned}&\sum_{\substack{m\le a/2\\ (m,a)=1}}\tfrac{2}{m}\log\tfrac{a/2}{m}+O(\log 2a)\\ =&\frac{\phi(a)}{a}\left((\log(a/2))^2-2(\log(a/2)\sum_{p|a}\frac{\log p}{p-1}\right)+O(\log 2a).\end{aligned}$$

When $\chi=\chi_0$, we have

$$\begin{aligned}f_\chi(s)&=\frac{L^3(s,\chi_0)}{L(2s,\chi_0)}\\ &=\frac{\zeta^3(s)\prod_{p|a}\left(1-\frac{1}{p^s}\right)^3}{\zeta(2s)\prod_{p|a}\left(1-\frac{1}{p^{2s}}\right)}\end{aligned}$$

$$= \frac{\zeta^3(s)}{\zeta(2s)} \prod_{p|a} \frac{(p^s-1)^2}{p^s(p^s+1)}.$$

Let

$$F(s) = \big((s-1)\zeta(s)\big)^3 \zeta(2s)^{-1} s^{-1},$$

$$G(s) = \prod_{p|a} \frac{(p^s-1)^2}{p^s(p^s+1)}$$

and

$$H(s) = F(s)G(s).$$

Then H has a removable singularity at $s = 1$ and we are concerned with the residue of

$$(s-1)^{-3} N^s H(s)$$

at $s = 1$. This is

$$\frac{1}{2} N(\log N)^2 H(1) + N(\log N)H'(1) + \frac{1}{2} N H''(1)$$

which it is convenient to rewrite as

$$NH(1)\left(\tfrac{1}{2}(\log N)^2 + (\log N)\tfrac{H'(1)}{H(1)} + \tfrac{H''(1)}{2H(1)}\right).$$

Now

$$\frac{H'(1)}{H(1)} = \frac{F'(1)}{F(1)} + \frac{G'(1)}{G(1)}$$

and

$$\frac{H''(1)}{H(1)} = \frac{F''(1)}{F(1)} + 2\frac{F'(1)G'(1)}{F(1)G(1)} + \frac{G''(1)}{G(1)}$$

and $F'(1)/F(1)$ and $F''(1)/F(1)$ can be evaluated in terms of Euler's and Stieltje's constants and $\zeta(2)$ and its derivatives. In particular

$$\tfrac{F'(1)}{F(1)} = 3\gamma - 2\tfrac{\zeta'(2)}{\zeta(2)} - 1.$$

The function G is more interesting. We have

$$\tfrac{G'(1)}{G(1)} = \textstyle\sum_{p|a} \tfrac{3p+1}{(p^2-1} \log p$$

and

$$\frac{G''(1)}{G(1)} = \left(\frac{G'(1)}{G(1)}\right)^2 - \sum_{p|a} \frac{3p^3+2p^2+3p}{(p^2-1)^2}(\log p)^2.$$

Thus

$$\frac{G'(1)}{G(1)} \ll \log\log(3a)$$

and

$$\frac{G''(1)}{G(1)} \ll (\log\log(3a))^2.$$

□

2.4 Proof of Theorem 2.2

By the same argument in the beginning of Section 2.3, $R(n;a)$ can be reduced to counting the number of divisors u of n^2 with $u + n \equiv 0 \pmod{a}$. Now the condition $u|n^2$ can be rewritten uniquely as $u = n_1 n_2^2$ and $n = n_1 n_2 n_3$ with n_1 being square-free. Thus we have

$$\begin{aligned} R(n;a) &= \sum_{\substack{u|n^2 \\ a|u+n}} 1 \\ &= \sum_{\substack{n_1 n_2 n_3 = n \\ a|n_2+n_3}} \mu^2(n_1) \end{aligned}$$

and hence

$$U(N) = \sum_{n_1 \le N} \mu^2(n_1) \left(\sum_{n_2 n_3 \le N/n_1} \sum_{\substack{a|n_2+n_3 \\ (a,n_1n_2n_3)=1}} 1 \right).$$

The inner double sum is symmetric in n_2 and n_3, so writing $M = N/n_1$ and using Dirichlet's method of the hyperbola it is

$$\sum_{n_2 \le \sqrt{M}} \sum_{\substack{a \le n_2 + M/n_2 \\ (a,n_1n_2)=1}} \sum_{\substack{n_3 \le M/n_2 \\ n_3 \equiv -n_2 \pmod{a}}} 2 - \sum_{n_2 \le \sqrt{M}} \sum_{\substack{a \le n_2 + \sqrt{M} \\ (a,n_1n_2)=1}} \sum_{\substack{n_3 \le \sqrt{M} \\ n_3 \equiv -n_2 \pmod{a}}} 1.$$

The second triple sum here is $\ll \sum_{n_2 \le \sqrt{M}} \sum_{a \le n_2 + \sqrt{M}} \frac{\sqrt{M}}{a} \ll M \log M$, leading to a contribution $\ll N(\log N)^2$ in the original sum. The first triple sum is

$$\sum_{n_2 \le \sqrt{M}} \sum_{\substack{a \le n_2 + M/n_2 \\ (a, n_1 n_2) = 1}} \frac{2M}{a n_2}$$

with an error $\ll M \log M$. The a in the range $(M/n_2, n_2 + M/n_2]$ are of order of magnitude M/n_2 and there are at most n_2 of them, so the total contribution from this part of the sum is $\ll M$, and the contribution from this to the original sum is $\ll N \log N$. Thus we are left with

$$\sum_{n_2 \le \sqrt{M}} \sum_{\substack{a \le M/n_2 \\ (a, n_1 n_2) = 1}} \frac{2M}{a n_2}.$$

Now using the Möbius function to pick out the condition $(a, n_1 n_2) = 1$, the inner sum over a can be written as

$$\sum_{k | n_1 n_2} \frac{\mu(k)}{k} \sum_{b \le M/(n_2 k)} \frac{2M}{b n_2}.$$

Put $k_1 = (k, n_1)$, $k_2 = k/k_1$, $n_1' = n_1/k_1$, so that $k_2 | n_2$, $(k_2, n_1') = 1$, and let $n_2' = n_2/k_2$. Observe also that for $\mu(n_1) = \mu(n_1' k_1) = \mu(n_1')\mu(k_1) \ne 0$ it is necessary that $(n_1', k_1) = 1$. Thus substituting in the original sum gives

$$\sum_{k_1 \le N} \sum_{k_2 \le N} \frac{\mu(k_1 k_2)}{k_1^2 k_2^2} \sum_{\substack{n_1' \le N/k_1 \\ (n_1', k_1 k_2) = 1}} \frac{\mu^2(n_1')}{n_1'} \sum_{n_2' \le k_2^{-1} \sqrt{N/(n_1' k_1)}} \sum_{b \le N/(n_1' n_2' k_1^2 k_2^2)} \frac{2N}{b n_2'}$$

and there are various implications for a non-zero contribution. Thus

$$n_1' n_2' k_1^2 k_2^2 \le N,$$

and this is a more stringent condition on n_2' than $n_2' \le k_2^{-1}\sqrt{N/(n_1' k_1)}$ when $n_2' \le k_1$. Also $n_1' \le N/(k_1^2 k_2^2)$ and $k_1 k_2 \le \sqrt{N}$. The sum over b is

$$\log\left(N/(n_1' n_2' k_1^2 k_2^2)\right) + O(1).$$

Consider the error term here. The sum over n_1' and n_2' contributes

$$\ll N(\log N)^2.$$

Thus one is left to consider

$$\sum_{\substack{k_1,k2\\k_1k_2\le\sqrt{N}}}\frac{\mu(k_1k_2)}{k_1^2k_2^2}\sum_{\substack{n_1'\le N/(k_1^2k_2^2)\\(n_1',k_2)=1}}\frac{\mu^2(n_1')}{n_1'}\sum_{\substack{n_2'\le k_2^{-1}\sqrt{N/(n_1'k_1)}\\n_2'\le N/(n_1'k_1^2k_2^2)}}\frac{2N}{n_2'}\log\left(\frac{N}{n_1'n_2'k_1^2k_2^2}\right).$$

The n_2' with $n_2'^2k_2^2n_1'k_1 \le N < n_2'n_1'k_1^3k_2^2$ satisfy $n_2' \le k_1$ so they would contribute $\ll N(\log k_1)\log N$ to the innermost sum and hence give a total contribution of $\ll N(\log N)^2$. Thus we can ignore the condition $n_2' \le N/(n_1'k_1^2k_2^2)$.

Now the the summation over n_2' can be performed and this gives

$$2N\left(\frac{1}{2}L_1^2+L_1L_2\right)$$

where

$$L_1=\log\frac{\sqrt{N}}{k_2\sqrt{n_1'k_1}}$$

and

$$L_2=\log\frac{\sqrt{N}}{k_1k_2\sqrt{n_1'k_1}}$$

with an error $\ll N\log N$ and a total error $\ll N(\log N)^2$. Now let

$$L=\log\frac{N}{k_1^2k_2^2n_1'},$$

then the above expression is easily seen to be a quadratic polynomial in L, *i.e.*

$$\begin{aligned}&2N\left(\frac{1}{2}L_1^2+L_1L_2\right)\\=\;&\frac{1}{2}\left(\frac{1}{2}(L+\log k_1)^2+(L+\log k_1)(L-\log k_1)\right)\\=\;&\frac{1}{4}\Big(3L^2+2(\log k_1)L-(\log k_1)^2\Big).\end{aligned}$$

Observe that the major contribution comes from the quadratic term in L here, and the other terms contribute $\ll N(\log N)^2$ in the original sum. So one is left to deal with

$$\frac{3}{4}\sum_{k_1\le\sqrt{N}}\sum_{k_2\le\sqrt{N}/k_1}\frac{\mu(k_1k_2)}{k_1^2k_2^2}\sum_{\substack{n_1'\le N/(k_1^2k_2^2)\\(n_1',k_1k_2)=1}}\frac{\mu^2(n_1')}{n_1'}N\left(\log\frac{N}{k_1^2k_2^2n_1'}\right)^2$$

$$=\frac{3}{4}\sum_{k\le\sqrt{N}}\frac{\mu(k)d(k)}{k^2}\sum_{\substack{n\le N/k^2\\(n,k)=1}}\frac{\mu^2(n)}{n}N\left(\log\frac{N}{k^2n}\right)^2.$$

When $\theta>0$ it follows by absolute convergence that the above sum is

$$\frac{3}{4\pi i}\int_{\theta-i\infty}^{\theta+i\infty}\zeta(1+s)D(1+s)\frac{N^{s+1}}{s^3}ds$$

where

$$D(s)=\prod_p\left(1-\frac{3}{p^{2s}}+\frac{2}{p^{3s}}\right).$$

The Euler product $D(s)$ converges locally uniformly for $\Re s>\frac{1}{2}+\delta$ for any $\delta>0$. Hence, by standard estimates for the Riemann zeta function the vertical path may be moved to the vertical path $\Re s=\psi$ where $-\frac{1}{2}<\psi<0$, picking up the residue of the pole of order 4 at $s=0$. It follows that

$$\frac{3}{4}\sum_{k\le\sqrt{N}}\frac{\mu(k)d(k)}{k^2}\sum_{\substack{n\le N/k^2\\(n,k)=1}}\frac{\mu^2(n)}{n}N\left(\log\frac{N}{k^2n}\right)^2$$

$$=\frac{3}{2}N\frac{(\log N)^3}{6}D(1)+O(N\log^2 N).$$

This establishes the theorem.

□

2.5 Further Comments

The corresponding questions for the equation (2.1) when $k\ge 3$ are still open. Indeed, whilst it follows from the criterion in the second paragraph of Section 2.3

that

$$R(n;a) \ll n^{\varepsilon},$$

and generally one could conjecture that $R_k(n;a)$, the number of solutions of (2.1) in positive integers, satisfies the concomitant bound

$$R_k(n;a) \ll n^{\varepsilon},$$

this is far from what has been established. Indeed, if we define $S_k(N;a)$ for general k by

$$S_k(N;a) = \sum_{\substack{n \le N \\ (n,a)=1}} R_k(n;a)$$

when $k \geq 3$ it has not even been established that

$$S_k(N;a) \ll N^{1+\varepsilon}.$$

It seems likely that

$$S_k(N;a) \sim CN(\log N)^{\alpha},$$

for some positive constants C and α which only depend on k and, in the case of C, on a. One can also make similar conjectures for the corresponding $T_k(N;a)$ and $U_k(N)$.

Chapter 3

Mean Value Theorems for Binary Egyptian Fractions II

3.1 Introduction

In Chapter 2 we studied the mean value

$$S(N;a) = \sum_{\substack{n \le N \\ (n,a)=1}} R(n;a), \tag{3.1}$$

of the number $R(n;a)$ of positive integer solutions to the Diophantine equation

$$\frac{a}{n} = \frac{1}{x} + \frac{1}{y}. \tag{3.2}$$

Here we extend our investigation to the second moment and some consequences thereof.

Theorem 3.1. *For fixed positive integer a, we have, for every $N \in \mathbb{N}$ with $N \geq 2$,*

$$\sum_{\substack{n \le N \\ (n,a)=1}} \left| R(n;a) - \frac{1}{\phi(a)} \sum_{\substack{\chi \bmod a \\ \chi^2 = \chi_0}} \bar{\chi}(-n) \sum_{u \mid n^2} \chi(u) \right|^2 \ll_a N \log^2 N,$$

where $\ll_a$ indicates that the implicit constant depends at most on a, and where χ_0 denotes the principal character modulo a.

In the character sum here the term $\chi = \chi_0$ contributes an amount $d(n^2)$ where d is the divisor function and we can expect that this is the dominant contribution on average. Thus as a consequence of the Erdős–Kac theorem, just as for the divisor function $d(n)$, one can anticipate that $\log R(n;a)$ admits a Gaussian distribution. As a first approximation we establish the normal order of $\log R(n;a)$.

Theorem 3.2. *When a is fixed, the normal order of $\log R(n;a)$ as a function of n is $(\log 3)\log\log n$.*

Let

$$\Phi(z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt.$$

Then with a little more work we can establish the full distribution.

Theorem 3.3. *For fixed positive integer a, we have*

$$\lim_{N\to\infty} \frac{1}{N} \mathrm{card} \left\{ \mathrm{n} \leq \mathrm{N} : \frac{\log \mathrm{R(n;a)} - (\log 3)\log\log \mathrm{n}}{(\log 3)\sqrt{\log\log \mathrm{n}}} \leq \mathrm{z} \right\} = \Phi(\mathrm{z}).$$

For completeness we also establish the mean square of $R(n;a)$ for fixed a. Since $R(n;a)$ resembles quite closely the divisor function $d(n^2)$ in many aspects, we expect that their mean squares share the same order of magnitude. Thus the following theorem can be compared with the asymptotic formula

$$\sum_{n\leq N} d^2(n^2) = NP_8(\log N) + O(N^{1-\delta})$$

which holds for some $\delta > 0$ and with $P_8(\cdot)$ a polynomial of degree 8.

Theorem 3.4. *Let a be a fixed positive integer and $\varepsilon > 0$. Then*

$$\sum_{\substack{n\leq N \\ (n,a)=1}} R(n;a)^2 = NP_8(\log N; a) + O_a(N^{35/54+\varepsilon})$$

where $P_8(\cdot\ ; a)$ is a degree 8 polynomial with coefficients depending on a, and its leading coefficient is

$$\frac{1}{8!a^2} \prod_{p|a} \left(1 - \frac{1}{p}\right)^7 \prod_{p\nmid a} \left(1 + \frac{6}{p} + \frac{1}{p^2}\right) \left(1 - \frac{1}{p}\right)^6.$$

The error term in the theorem above is closely related to the generalised divisor problem, and in particular depends on a mean value estimate for the ninth moment of Dirichlet L-functions $L(s,\chi)$ modulo a inside the critical strip. As is easily verified, the error can be improved to $O_a(N^{1/2+\varepsilon})$ under the assumption of the generalised Lindelöf Hypothesis.

3.2 Proof of Theorem 3.1

We rewrite equation (3.2) in the form

$$(ax-n)(ay-n)=n^2.$$

After the change of variables $u=ax-n$ and $v=ay-n$, it follows that $R(n;a)$ is the number of ordered pairs of natural numbers u, v such that $uv=n^2$ and $u\equiv v\equiv -n \pmod a$.

Under the assumption that $(n,a)=1$, $R(n;a)$ can be reduced further to counting the number of divisors u of n^2 with $u\equiv -n \pmod a$. Now the residue class $u\equiv -n \pmod a$ is readily isolated *via* the orthogonality of the Dirichlet characters χ modulo a. Thus we have

$$R(n;a)=\frac{1}{\phi(a)}\sum_{\substack{\chi\\ \bmod a}}\bar{\chi}(-n)\sum_{u|n^2}\chi(u), \tag{3.3}$$

where the condition $(n,a)=1$ is taken care of by the character $\bar{\chi}(n)$.

Hence

$$\begin{aligned}
&\sum_{\substack{n\le N\\ (a,n)=1}}\left|R(n;a)-\frac{1}{\phi(a)}\sum_{\substack{\chi \bmod a\\ \chi^2=\chi_0}}\bar{\chi}(-n)\sum_{u|n^2}\chi(u)\right|^2\\
\ll_a\;&\sum_{\substack{n=1\\ (a,n)=1}}^{\infty}e^{-n/N}\left|\sum_{\substack{\chi \bmod a\\ \chi^2\neq\chi_0}}\bar{\chi}(-n)\sum_{u|n^2}\chi(u)\right|^2\\
=\;&\sum_{\substack{\chi_1 \bmod a\\ \chi_1^2\neq\chi_0}}\sum_{\substack{\chi_2 \bmod a\\ \chi_2^2\neq\chi_0}}\bar{\chi}_1\chi_2(-1)\sum_{n=1}^{\infty}\sum_{u|n^2}\sum_{v|n^2}\chi_1(u)\bar{\chi}_2(v)\bar{\chi}_1\chi_2(n)e^{-n/N},
\end{aligned}$$

where χ_0 denotes the principal character modulo a. In order to evaluate the sum over n, we analyze the Dirichlet series

$$f_{\chi_1,\chi_2}(s) := \sum_{n=1}^{\infty}\sum_{u|n^2}\sum_{v|n^2}\chi_1(u)\bar{\chi}_2(v)\bar{\chi}_1\chi_2(n)n^{-s}.$$

The condition $u|n^2$ can be written as $u_1u_2^2|n^2$ with u_1 squarefree, i.e. $u_1u_2|n$, and likewise for $v|n^2$. Thus

$$f_{\chi_1,\chi_2}(s) = \sum_{m=1}^{\infty}\frac{\bar{\chi}_1\chi_2(m)}{m^s}\sum_{d=1}^{\infty}\frac{F(d)}{d^s} \tag{3.4}$$

where

$$F(d) = \sum_{\substack{u_1,u_2,v_1,v_2\\ d=[u_1u_2,v_1v_2]}}\mu^2(u_1)\mu^2(v_1)\chi_1(u_1u_2^2)\bar{\chi}_2(v_1v_2^2)\bar{\chi}_1\chi_2(d).$$

The function F is multiplicative and so the inner sum above is

$$\prod_p\left(1+\sum_{k=1}^{\infty}F(p^k)p^{-ks}\right), \tag{3.5}$$

where

$$F(p^k) = \sum_{\substack{u_1,u_2,v_1,v_2\\ [u_1u_2,v_1v_2]=p^k}}\mu^2(u_1)\mu^2(v_1)\chi_1(u_1u_2^2)\bar{\chi}_2(v_1v_2^2)\bar{\chi}_1\chi_2(p^k). \tag{3.6}$$

In particular we have

$$F(p) = \chi_0(p) + \sum_{\chi\in\mathcal{X}\setminus\{\bar{\chi}_1\chi_2\}}\chi(p),$$

where $\mathcal{X} = \{\chi_1,\chi_2,\chi_1\chi_2,\chi_1\bar{\chi}_2,\bar{\chi}_1,\bar{\chi}_2,\bar{\chi}_1\bar{\chi}_2,\bar{\chi}_1\chi_2\}$ (and the entries are considered to be formally distinct), and

$$|F(p^k)| \le 8k.$$

Thus the Dirichlet series f converges absolutely for $\sigma > 1$ and

$$f_{\chi_1,\chi_2}(s) = G_{\chi_1,\chi_2}(s)L(s,\chi_0)\prod_{\chi\in\mathcal{X}}L(s,\chi), \tag{3.7}$$

where $G_{\chi_1,\chi_2}(s)$ is a function which is analytic in the region $\Re s > 1/2$ and satisfies

$$G(s) \ll 1 \quad (\sigma \geq \tfrac{1}{2} + \delta)$$

for any fixed $\delta > 0$. As χ_1, χ_2 are not characters of order 1 or 2, $f_{\chi_1,\chi_2}(s)$ has a triple pole at $s = 1$ when $\chi_1 = \chi_2$ or $\chi_1\chi_2 = \chi_0$, and a simple pole otherwise. By Corollary 1.17 and Lemma 10.15 of [28], for fixed a,

$$L(s,\chi) - \frac{E(\chi)\phi(a)}{a(s-1)} \ll 2 + |t|$$

uniformly for $\sigma \geq \frac{1}{2}$ where $E(\chi)$ is 1 when $\chi = \chi_0$ and 0 otherwise. Hence by (5.25) of [28]

$$\sum_{n=1}^{\infty}\sum_{u|n^2}\sum_{v|n^2}\chi_1(u)\bar{\chi}_2(v)\bar{\chi}_1\chi_2(n)e^{-n/N} = \frac{1}{2\pi i}\int_{\theta-i\infty}^{\theta+i\infty} f_{\chi_1,\chi_2}(s)N^s\Gamma(s)ds,$$

where $\theta > 1$. Since the gamma function decays exponentially fast on any vertical line we may move the vertical path to the $\frac{3}{4}$–line picking up the residue of the integrand at $s = 1$. The residue contributes an amount

$$\ll N(\log N)^2$$

and the new vertical path contributes

$$\ll N^{\frac{3}{4}}.$$

This completes the proof of Theorem 3.1.

3.3 Proof of Theorem 3.2

By Theorem 3.1, we expect that for almost all n with $(a,n) = 1$, $R(n;a)$ is close to

$$\frac{1}{\phi(a)}\sum_{\substack{\chi \bmod a\\ \chi^2=\chi_0}}\bar{\chi}(-n)\sum_{u|n^2}\chi(u).$$

Thus we need to examine the contribution from the characters modulo a of order 1 and 2. For general a, there may be many quadratic characters modulo a. Nevertheless we believe that the major contribution to the sum above comes from the principal character, and this is of size

$$\frac{d(n^2)}{\phi(a)}.$$

Thus, for fixed a, $\log R(n;a)$ should have the normal order of $\log d(n^2)$, namely $(\log 3)\log\log n$. When $(n,a) > 1$ we have

$$R(n;a) = R(n/(n,a); a/(n,a)) \tag{3.8}$$

and so we can expect that the general case follows from the special case $(n,a) = 1$.

Before embarking on the proof of Theorem 3.2, we state a lemma. We define, for any quadratic character χ,

$$\Omega_\chi(n) = \text{card}\left\{p, k : k \geq 1, p^k|n, \chi(p^k) = 1\right\}.$$

Lemma 3.5. *Suppose that χ is a quadratic character to a fixed modulus a and that $N \geq 3$. Then*

$$\sum_{n\leq N}\left(\Omega_\chi(n) - \frac{1}{2}\log\log N\right)^2 \ll N\log\log N$$

and

$$\sum_{1<n\leq N}\left(\Omega_\chi(n) - \frac{1}{2}\log\log n\right)^2 \ll N\log\log N.$$

Proof. The proof follows in the same way as Turán's theorem (see Theorem 2.12 of [28]) on observing that

$$\sum_{\substack{p\leq N\\ \chi(p)=1}}\frac{1}{p} = \frac{1}{2}\log\log N + O(1)$$

and this is readily deduced from Corollary 11.18 of [28]. □

It is an immediate consequence of the above lemma that $\Omega_\chi(n)$ has normal order $\frac{1}{2}\log\log n$. In particular, for any fixed $\varepsilon > 0$, for almost all n,

$$3^{\Omega_\chi(n)} < 3^{(\frac{1}{2}+\varepsilon)\log\log n}.$$

Now, for any quadratic character χ modulo a, let

$$g_\chi(n) = \sum_{u|n^2} \chi(u).$$

This is

$$\prod_{p^k\|n} \left(1 + \chi(p) + \chi^2(p) + \cdots + \chi^{2k}(p)\right).$$

When $\chi(p) = -1$ the general factor is 1, and when $\chi(p) = 1$ it is $2k+1$. Hence

$$0 < g_\chi(n) \le 3^{\Omega_\chi(n)}.$$

Thus for any fixed $\varepsilon > 0$, for every quadratic character modulo a, for almost all n,

$$g_\chi(n) < (\log n)^{(\frac{1}{2}\log 3+\varepsilon)}. \tag{3.9}$$

Let

$$r(n;a) = \frac{1}{\phi(a/(n,a))} \sum_{\substack{\chi \bmod a/(n,a) \\ \chi^2=\chi_0}} \bar{\chi}(-n/(n,a)) g_\chi(n/(n,a)).$$

Since $R(n;a) = R(n/(n,a); a/(n,a))$, it follows by Theorem 3.1 that

$$\sum_{n\le N} (R(n;a) - r(n;a))^2 = \sum_{d|a} \sum_{\substack{m\le N/d \\ (m,a/d)=1}} (R(m;\tfrac{a}{d}) - r(m;\tfrac{a}{d}))^2 \ll N(\log N)^2.$$

Hence, for any fixed $\varepsilon > 0$, for almost all n we have

$$|R(n;a) - r(n;a)| < (\log n)^{1+\varepsilon}.$$

Therefore, by (3.9), for almost all n,

$$\left| R(n;a) - \frac{d\left((n/(a,n))^2\right)}{\phi(a/(a,n))} \right| < (\log n)^{1+2\varepsilon}. \tag{3.10}$$

Now $3 \le d(p^{2k}) = 1 + 2k \le 3^k$. Hence

$$3^{\omega(n)-\omega(a)} \le d\left((n/(a,n))^2\right) \le 3^{\Omega(n)} \tag{3.11}$$

and it follows that

$$(\log n)^{\log 3-\varepsilon} < \frac{d\left((n/(a,n))^2\right)}{\phi(a/(a,n))} < (\log n)^{\log 3+\varepsilon}$$

for almost all n. Theorem 3.2 now follows.

3.4 Proof of Theorem 3.3

By (3.10) and (3.11), for every fixed $\varepsilon > 0$, for almost all n,

$$\frac{3^{\omega(n)}}{\phi(a/(a,n))} - (\log n)^{1+\varepsilon} < R(n;a) < 3^{\Omega(n)} + (\log n)^{1+\varepsilon}.$$

Moreover, for almost all n we have $\Omega(n) \ge \omega(n) > (1-\varepsilon)\log\log n$. Hence for any δ with $0 < \delta < \log 3 - 1$ we have, for almost all n

$$3^{\omega(n)-\omega(a)-\log\phi(a/(a,n))}\exp(-(\log n)^{-\delta}) < R(n;a) < 3^{\Omega(n)}\exp((\log n)^{-\delta})$$

and so

$$3^{\omega(n)}\exp(-\varepsilon\sqrt{\log\log n}) < R(n;a) < 3^{\Omega(n)}\exp(\varepsilon\sqrt{\log\log n}).$$

Let

$$S(N;z) = \mathrm{card}\left\{ n \le N : \frac{\log R(n;a) - (\log 3)\log\log n}{\log 3\sqrt{\log\log n}} \le z \right\},$$

$$S_{-}(N;z) = \mathrm{card}\left\{ n \le N : \frac{\Omega(n) - \log\log n}{\sqrt{\log\log n}} \le z \right\}$$

and

$$S_+(N;z) = \text{card}\left\{n \le N : \frac{\omega(n) - \log\log n}{\sqrt{\log\log n}} \le z\right\},$$

Then for a non-negative monotonic function $\eta(n)$ tending to 0 sufficiently slowly as $N \to \infty$ we have

$$-\eta(N)N + S_-(N; z-\varepsilon) < S(N;z) < \eta(N)N + S_+(N; z+\varepsilon).$$

Hence, by the Erdős–Kac theorem (see, for example Theorem 7.21 and Exercise 7.4.4 of [28]),

$$\Phi(z-\varepsilon) \le \liminf_{N\to\infty} N^{-1}S(N;z) \le \limsup_{N\to\infty} N^{-1}S(N;z) \le \Phi(z+\varepsilon).$$

The theorem now follows from the continuity of Φ.

3.5 Proof of Theorem 3.4

By a similar discussion to that in Section 3.2, we can show that the generating Dirichlet series for $R(n;a)^2$ is

$$\sum_{\substack{n=1 \\ (n,a)=1}}^{\infty} \frac{R(n;a)^2}{n^s} = \frac{1}{\phi(a)^2} \sum_{\substack{\chi_1,\chi_2 \\ \text{mod } a}} \bar{\chi}_1\chi_2(-1) f_{\chi_1,\chi_2}(s),$$

where $f_{\chi_1,\chi_2}(s)$ is analytic in the region $\Re s > 1/2$ and is given by (3.7). Here $f_{\chi_1,\chi_2}(s)$ has a pole at 1 of order at least 1, and as high as 9 exactly when χ_1 and χ_2 are equal to the principal character χ_0. Now on applying Perron's formula, we have for $\theta = 1 + 1/\log(2N)$,

$$\sum_{\substack{n\le N \\ (n,a)=1}} R(n;a)^2 = \sum_{\substack{\chi_1,\chi_2 \\ \text{mod } a}} \frac{\bar{\chi}_1\chi_2(-1)}{2\pi i\phi(a)^2} \int_{\theta-iT}^{\theta+iT} f_{\chi_1,\chi_2}(s)\frac{N^s}{s}ds + O_a(N^{1+\varepsilon}/T). \quad (3.12)$$

Since we are shooting for the asymptotics for the the mean square, smoothing factors of the kind used in Section 3.2 are best avoided. Since the integrand includes a product of nine L–functions, we cannot expect to be able to move the vertical integral path too close to the 1/2-line in the current state of knowledge. Nevertheless, the following result of Meurman [24] which extends Heath-Brown's theorem [11] on the twelfth power moment of the Riemann zeta function to Dirichlet L–functions, provides a starting point for the analysis.

Lemma 3.6.
$$\sum_{\chi \bmod a} \int_{-T}^{T} |L(\tfrac{1}{2}+it,\chi)|^{12} dt \ll a^3 T^{2+\varepsilon},$$
where $\varepsilon > 0$, $a \geq 1$ *and* $T \geq 2$.

Then, adapting the argument of Chapter 8 of Ivić [21] for the Riemann zeta function to Dirichlet L-functions establishes the following.

Lemma 3.7.
$$\int_{-T}^{T} |L(\tfrac{35}{54}+it,\chi)|^{9} dt \ll_a T^{1+\varepsilon},$$
where $\varepsilon > 0$, χ *is a fixed Dirichlet character modulo* $a \geq 1$ *and* $T \geq 2$.

If one utilizes the sharpest estimates for the underlying exponential sums, Lemma 3.7 is susceptible to slight improvements.

Now, we move the vertical integral path in (3.12) to the 35/54-line, picking up the residue of the integrand at 1. Thus

$$\int_{\theta-iT}^{\theta+iT} f_{\chi_1,\chi_2}(s)\frac{N^s}{s}ds = \int_{\theta-iT}^{35/54-iT} f_{\chi_1,\chi_2}(s)\frac{N^s}{s}ds + \int_{35/54+iT}^{\theta+iT} f_{\chi_1,\chi_2}(s)\frac{N^s}{s}ds + \int_{35/54-iT}^{35/54+iT} f_{\chi_1,\chi_2}(s)\frac{N^s}{s}ds + \mathrm{Res}_{s=1}\left(f_{\chi_1,\chi_2}(s)\frac{N^s}{s}\right)$$

Here, in order to deal with the contribution from the horizontal integrals, we cannot afford to use the crude convexity bounds on Dirichlet L-functions, due to the large number of L–functions in the integrand. Fortunately, a sharper bound has been established by C.D. Pan and C.B. Pan in Theorem 24.2.1 of [30].

Lemma 3.8. *Let* $l \geq 3$, $L = 2^{l-1}$ *and* $\sigma_l = 1 - l(2L-2)^{-1}$. *Then when* $\sigma \geq \sigma_l$

$$L(\sigma+it,\chi) \ll_a |t|^{1/(2L-2)} \log|t|$$

holds uniformly for $|t| \geq 2$.

When $l = 3$ we obtain

$$L(\sigma + it, \chi) \ll_a |t|^{1/6} \log |t|$$

uniformly for $|t| \geq 2$ and $\sigma \geq \frac{1}{2}$, and when $l = 4$,

$$L(\sigma + it, \chi) \ll_a |t|^{1/14} \log |t|$$

uniformly for $\sigma \geq 5/7$. Thus, by the convexity principle for Dirichlet series,

$$L(\sigma + it, \chi) \ll_a |t|^{\mu(\sigma)+\varepsilon}$$

uniformly for $|t| \geq 2$ and $\sigma \geq \frac{1}{2}$ where

$$\mu(\sigma) = \begin{cases} \frac{1}{6} - \frac{4}{9}(\sigma - \frac{1}{2}) & \text{when } \frac{1}{2} \leq \sigma \leq \frac{5}{7}, \\ \frac{1-\sigma}{4} & \text{when } \frac{5}{7} < \sigma \leq 1, \\ 0 & \text{when } 1 < \sigma. \end{cases}$$

We note that $\mu(\frac{35}{54}) = \frac{49}{486} < \frac{1}{9}$ and $\mu(\frac{5}{7}) = \frac{1}{14}$.

Now the horizontal paths contribute

$$\ll \int_{35/54}^{1+\varepsilon} N^{\sigma} |f_{\chi_1,\chi_2}(\sigma + iT)| T^{-1} d\sigma.$$

and this is

$$\ll \max_{35/54 \leq \sigma \leq 1+\varepsilon} N^{\sigma} T^{9\mu(\sigma)-1+\varepsilon},$$

and by the piecewise linearity of σ and $\mu(\sigma)$ this is

$$\ll N^{1+\varepsilon} T^{-1} + N^{5/7} T^{9\mu(5/7)-1+\varepsilon} + N^{35/54} T^{9\mu(35/54)-1+\varepsilon}.$$

When $T = N$ this is

$$\ll N^{35/54+\varepsilon}.$$

On the other hand, by Lemma 3.7 the vertical path also contributes

$$\ll N^{35/54+\varepsilon}.$$

The main term comes from the residual contributions, which, in the case that $\chi_1 = \chi_2 = \chi_0$, is $NP_8(\log N; a)$ where $P_8(\cdot\,; a)$ is a polynomial of degree 8 whose coefficients depend on a. Notice that for other choices of χ_1 and χ_2, the residual contribution gives a polynomial of $\log N$ of lower degree than above.

For the leading coefficient, we need more precise information about f_{χ_0,χ_0}. By (3.4), (3.5) and (3.6) we have

$$f_{\chi_0,\chi_0} = L(s,\chi_0)\prod_{p\nmid a}\left(1+\sum_{k=1}^{\infty}\frac{8k}{p^{ks}}\right) \tag{3.13}$$

$$= L(s,\chi_0)^9\prod_{p\nmid a}(1+6p^{-s}+p^{-2s})(1-p^{-s})^6, \tag{3.14}$$

from which the leading coefficient is readily deduced. This completes the proof of Theorem 3.4.

In conclusion we remark that a concomitant argument will give

$$\sum_{\substack{n\leq N\\(n,a)=1}} R(n;a)^k = NP_{3^k-1}(\log N; a) + O_a\left(N^{\alpha_k+\varepsilon}\right)$$

for any $\varepsilon > 0$, where $P_{3^k-1}(\cdot\,; a)$ is a polynomial of degree 3^k-1 whose coefficients depend on a and α_k is a constant that depends on the best 3^k-th power moment estimates for $L(s,\chi)$ in the critical strip and the quantity $\mu(\sigma)$ defined above. This question is closely related to the generalised divisor problem, and one is referred to Chapter 13 in Ivić [21] for more details.

Chapter 4

On the Exceptional Set for Binary Egyptian Fractions

4.1 Introduction

Let a be a fixed positive integer. We consider the binary Diophantine equation

$$\frac{a}{n} = \frac{1}{x} + \frac{1}{y} \tag{4.1}$$

and denote by $R(n;a)$ the number of pairs of positive integer solutions (x,y) satisfying the equation (4.1). A good deal is now known about the average behaviour of $R(n;a)$. See [2], [17] and [18] for details. In this chapter, we are concerned with the number of n such that the equation (4.1) is not soluble in positive integers x and y and to this end we define

$$\mathcal{E}_a = \{n \in \mathbb{N} : R(n;a) = 0\}.$$

Clearly both $\mathcal{E}_1$ and $\mathcal{E}_2$ are empty. When $a \geq 3$ the structure of $\mathcal{E}_a$ is more delicate and of great interest. In this paper, we investigate the asymptotic size of $\mathcal{E}_a$. Thus we define

$$\mathcal{E}_a(N) = \{n \in \mathcal{E}_a : n \leq N\}$$

and

$$E_a(N) = \#\mathcal{E}_a(N).$$

In 1985, G. Hofmeister and P. Stoll [13] proved that the set $\mathcal{E}_a$ has asymptotic density 0, and more precisely that

$$E_a(N) \ll_a \frac{N}{(\log N)^{1/\phi(a)}}.$$

For $a = 5$ and $a \geq 7$ this bound is far from the truth. Their method is based on the observation that if the equation (4.1) is insoluble, then n is not divisible by any prime of the form $p \equiv -1 \pmod{a}$. Thus a simple application of Selberg's upper bound sieve gives the stated bound. However when $a = 5$ or $a \geq 7$ the bulk of the n deficient in such prime factors nevertheless have a representation.

The following theorem establishes the precise asymptotic behaviour of $E_a(N)$.

Theorem 4.1. *For fixed $a \geq 3$, let $2^{\gamma_0} p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k}$ be the canonical decomposition of a and define m and δ by*

$$2^m \| \gcd(\delta, p_1 - 1, p_2 - 1, \cdots, p_k - 1)$$

and

$$\delta = \begin{cases} 0, & \text{if } \gamma_0 \leq 1, \\ 2, & \text{if } \gamma_0 \geq 2. \end{cases}$$

Then we have

$$E_a(N) \sim C(a) \frac{N (\log\log N)^{2^{m-1}-1}}{(\log N)^{1-1/2^m}}$$

where $C(a)$ is a positive constant depending only on a.

In order to establish this theorem we need first to investigate the underlying structure of $\mathcal{E}_a$, and we embark on this in Section 4.2. The case when a is a prime power is somewhat easier to understand and, having established some preliminary lemmata in 4.2.1, we consider this case in Section 4.2.2. This then leads into a discussion of the general case in 4.2.3.

In Section 4.3 the main analytic input is introduced, and it is convenient to base this on an arithmetical application of a theorem of Delange. Delange's theorem is a refinement of the Wiener–Ikehara theorem and is qualitative in nature. In particular it does not give an explicit error term. By using instead a method allied to that leading to the strongest known unconditional error term in the prime number theorem it would be possible to give a quantitative error term in Lemma 4.11 of a similar quality. However whilst this would be quite routine in nature there would be many detailed complications and more importantly the extra effort would not lead to any further illumination of the central problem of this paper in that a greater loss in the error term appears at a later stage of our argument. We are happy to leave this approach as an exercise to the reader.

The proof of the main theorem is completed in Section 4.4 through a suitable combination of Sections 4.2 and 4.3.

After we submitted our paper [19] to arXiv, C. Elsholtz pointed out his thesis [5] and a subsequent note to us, in which he adumbrates the special case of Theorem 4.1 that a is a prime. In particular, in his thesis it is showed that $E_a(N) \ll_a N/\sqrt{\log N}$ for all $a \geq 1$ and that $E_a(N) \gg_a N/\sqrt{\log N}$ for $a \equiv 0, 3 \pmod 4$. Furthermore, the asymptotic formula for the particular case $a = 13$ has been worked out in his unpublished work. However, no strategy towards the proof of the full result (even for the case that a is a prime) was present in either his thesis or the unpublished note.

Throughout this chapter, we reserve the letters p, q and r for prime numbers and calligraphic letters for sets and sequences. In particular, if $\mathcal{A} \subseteq \mathbb{N}$ we denote by $\mathcal{A}(N)$ the subset of $\mathcal{A}$ with elements less than or equal to N and $|\mathcal{A}(N)|$ denotes the cardinality of $\mathcal{A}(N)$.

4.2 The Structure of $\mathcal{E}_a$

4.2.1 Some elementary lemmata

It is more convenient to work with the notations $\mathcal{E}_a^*$, $\mathcal{E}_a^*(N)$ and $E_a^*(N)$, defined as follows:

$$\mathcal{E}_a^* = \{n \in \mathcal{E}_a : (n, a) = 1\}$$

and $\mathcal{E}_a^*(N)$ and $E_a^*(N)$ can be defined accordingly. Then we have immediately the following.

Lemma 4.2. *We have*

$$\mathcal{E}_a(N) = \bigcup_{d|a} \mathcal{E}_{a/d}^*(N/d)$$

and hence have

$$E_a(N) = \sum_{d|a} E^*_{a/d}(N/d).$$

Thus the structure of $\mathcal{E}_a$ can be deduced readily from that of $\mathcal{E}^*_a$. Henceforward we assume that n is a positive integer coprime to a, unless otherwise stated.

The starting point of our argument is the following elementary lemma.

Lemma 4.3. *The equation* (4.1) *with* $(a, n) = 1$ *is soluble in positive integers if and only if there exists a pair of coprime factors* u *and* v *of* n *such that* $a|u + v$*.*

Proof. If the equation (4.1) is soluble, then we rewrite it as $axy = n(x + y)$, let $(x, y) = l$ and write $x = ul$ and $y = vl$ with (u,v)=1. Thus $aluv = n(u+v)$. Then, as $(a, n) = 1$ and $(uv, u + v) = 1$, we have $uv|n$ and $a|u + v$.

In the opposite direction, we write $u + v = aa'$ and $n = uvn'$, so that $\frac{a}{n} = \frac{1}{a'n'u} + \frac{1}{a'n'v}$. □

This lemma suggests that the solubility of equation (4.1) solely depends on the residue classes of factors of n modulo a, and hence depends on the residue classes of prime factors of n modulo a, which naturally leads our discussion to the distribution of prime factors of n in the multiplicative group $(\mathbb{Z}/a\mathbb{Z})^*$.

4.2.2 The case that a is a power of odd prime

We consider the case that $a = p^\gamma$ is a power of odd prime in this subsection and come back to the general case later. This strategy fits with both the motivational purpose and the presentational purpose. Let G denote the cyclic group $(\mathbb{Z}/a\mathbb{Z})^*$ of reduced residue classes modulo $a = p^\gamma$, and let H be the maximal subgroup of $G = (\mathbb{Z}/a\mathbb{Z})^*$ with cardinality $|H|$ being odd, namely H is the maximal subgroup of G such that $\overline{-1} \notin H$ and clearly such a group is unique. Here and throughout this article $\bar{i}$ means the residue class $i \pmod a$, if there is no ambiguity about the modulus a in the context. Now let $\phi(a) = 2^m d$ with d being an odd number. If we fix a primitive root g modulo a, then

$$G = \{g, g^2, g^3, \ldots, g^{2^m d}\} \tag{4.2}$$

and

$$H = \{g^{2^m}, g^{2\cdot 2^m}, g^{3\cdot 2^m}, \ldots, g^{d\cdot 2^m}\}, \tag{4.3}$$

by which one readily verifies that $\overline{-1} \notin H$ since $g^{\phi(a)/2} \equiv -1 \pmod{a}$. Hence we have the index $[G : H] = 2^m$ and $|H| = \frac{\phi(a)}{2^m} = d$.

Essentially the structure of $\mathcal{E}_a^*$ is that any $n \in \mathcal{E}_a^*$ can have arbitrarily many prime factors lying in the residue classes in H but can have at most a bounded number of prime factors lying outside H. It is this observation that renders the counting function of $\mathcal{E}_a^*$ susceptible to an analytic argument.

Lemma 4.4. *We have the following inclusion relation of sets*

$$\{n \in \mathbb{N} : q|n \text{ with } q \text{ being prime } \Rightarrow \overline{q} \in H\} \subseteq \mathcal{E}_a^*.$$

Proof. For any n on the left hand side, and for any pair of coprime positive integers u and v with $uv|n$ we have $\overline{u}, \overline{v} \in H$ in light of the fact that H is a group. Since $\overline{-1} \notin H$, we have $\overline{-v} \notin H$ and hence $\overline{u} \neq \overline{-v}$, in other words $a \nmid u + v$. Now Lemma 4.4 follows from Lemma 4.3. □

The next lemma is central to our understanding of the structure of $\mathcal{E}_a^*$.

Lemma 4.5. *Let $m \geq 1$ and $\mathcal{G}$ denote the additive group $\mathbb{Z}/(2^m\mathbb{Z})$, let $\{e_j\}_1^t$ be a sequence with t nonzero elements of $\mathcal{G}$ (i.e., repeated elements are allowed in $\{e_j\}$), and form the subset of $\mathcal{G}$*

$$\mathcal{S} = \left\{ \sum_{j=1}^{t} \delta_j e_j : \delta_j \in \{-1, 0, 1\} \right\}.$$

(i) *If $t \geq 2^{m-1}$, then $\overline{2^{m-1}} \in \mathcal{S}$. Namely, as long as the length of $\{e_j\}$ exceeds 2^{m-1}, for whatever choices of the elements e_j, one can always find a partial sum, as in the definition of $\mathcal{S}$, such that it is equal to $\overline{2^{m-1}}$.*

(ii) *If $t = 2^{m-1} - 1$, then the corresponding set $\mathcal{S}$ does not contain $\overline{2^{m-1}}$ if and only if the sequence $\{e_j\}_1^t$ satisfies $e_j \equiv \pm e \pmod{2^m}$ for each j and some fixed $e \in (\mathbb{Z}/(2^m)\mathbb{Z})^*$.*

Proof. We prove (i) first. Note that if one of the e_j satisfies $e_j \equiv 0 \pmod{2^{m-1}}$ then (i) is automatically true by choosing the δ_j for that particular j to be 1 and all the others to be 0. So without loss of generality we assume none of the e_j for $1 \le j \le t$ is divisible by 2^{m-1}. The proof is by induction on m. The initial case $m = 1$ is trivial. Thus we can suppose that $m \ge 1$ and that the conclusion is true for m. Consider $\mathcal{G} = \mathbb{Z}/(2^{m+1}\mathbb{Z})$ and a sequence $\{e_j\}_1^{2^m} \subseteq \mathcal{G}$. Note that none of the e_j is congruent to 0 modulo 2^m and hence by the induction assumption, we know that there exist $\delta_j \in \{-1, 0, 1\}$ for $1 \le j \le 2^m$ such that

$$s_1 := \sum_{j=1}^{2^{m-1}} \delta_j e_j \equiv 2^{m-1} \pmod{2^m}$$

and

$$s_2 := \sum_{j=2^{m-1}+1}^{2^m} \delta_j e_j \equiv 2^{m-1} \pmod{2^m}.$$

Choose u_i so that $s_i = 2^{m-1} + u_i 2^m$ for $i \in \{1, 2\}$. Then by considering separately the cases when the u_i are of the same or differing parity it follows that either $s_1 + s_2$ or $s_1 - s_2$ is congruent to 2^m modulo 2^{m+1}. This establishes (i). The proof of (ii) is similar but a little more elaborate. If there is an $e \in (\mathbb{Z}/(2^m)\mathbb{Z})^*$ such that $e_j \equiv \pm e \pmod{2^m}$ for every j, then regardless of the choice of δ_j we have $\sum_{j=1}^{t} \delta_j e_j \equiv \pm ue \pmod{2^m}$ where $|u| \le 2^{m-1} - 1$. Thus $\overline{2^{m-1}} \notin \mathcal{S}$. Thus it remains to consider the situation when $\overline{2^{m-1}}$ is not contained in $\mathcal{S}$. As before, we argue by induction on m. When $m = 1$ we have $t = 0$ and $\mathcal{S}$ is empty so the conclusion is trivial. When $m = 2$ we have $t = 1$ and $2^{m-1} = 2$, and so $e_1 \not\equiv 0$ or $2 \pmod 4$ and we are done. Now suppose that the conclusion holds for a given value of $m \ge 2$ and consider the case with m replaced by $m + 1$. That is, we suppose that $\overline{2^m}$ is not contained in $\mathcal{S}$ and will deduce that there is an $e \in (\mathbb{Z}/(2^{m+1})\mathbb{Z})^*$ such that each e_j satisfies $e_j \equiv \pm e \pmod{2^{m+1}}$. We now form the partial sums

$$s_1 := \sum_{j=1}^{2^{m-1}-1} \delta_j e_j$$

and

$$s_2 := \sum_{j=2^{m-1}}^{2^m-1} \delta_j e_j.$$

By (i) and the inductive hypothesis if there is no e such that $e_j \equiv \pm e \pmod{2^m}$

for $1 \leq j \leq 2^{m-1} - 1$, where $e \in (\mathbb{Z}/(2^m)\mathbb{Z})^*$, then there is a choice of the δ_j such that

$$s_2 \equiv 2^{m-1} \pmod{2^m}$$

and

$$s_1 \equiv 2^{m-1} \pmod{2^m}.$$

Thus if there is no such e, then as before one of $s_1 \pm s_2 \equiv 2^m \pmod{2^{m+1}}$, which we have expressly excluded. Thus there is such an e. Moreover we can repeat the argument with every permutation of the e_j. Thus we can conclude that there is an e such that $e_j \equiv \pm e \pmod{2^m}$ for $1 \leq j \leq 2^m - 1$, where $e \in (\mathbb{Z}/(2^m)\mathbb{Z})^*$. In other words

$$e_j \equiv \pm e \text{ or } \pm (e + 2^m) \pmod{2^{m+1}}.$$

Now we may conclude that either all the e_j are congruent to $\pm e$ or they are congruent to $\pm(e + 2^m)$, because if, say, $e_1 \equiv \pm e \pmod{2^{m+1}}$ and $e_2 \equiv \pm(e + 2^m) \pmod{2^{m+1}}$ then either $e_1 + e_2$ or $e_1 - e_2$ is $2^m \pmod{2^{m+1}}$, contradicting $\overline{2^m} \notin \mathcal{S}$. □

A weaker version of the lemma in which one replaces the exact lower bound 2^{m-1} of t in part (i) by the crude bound $(2^m - 1)(2^{m-1} - 1) + 1$, would follow by a direct application of the pigeonhole principle. An extension of this lemma to general modulus (not necessarily a power of 2) could be formulated and then proved by Kneser's theorem (see chapter 1 in [9]), which is, however, not of direct relevance to the purpose of this memoir. Nevertheless, it would be of the essence if one desires to establish the lower order terms for the asymptotics in Theorem 4.1.

Having established the necessary preliminaries, we are poised to reveal the structure of $\mathcal{E}_a^*$ when $a = p^\gamma$ is a power of an odd prime.

Lemma 4.6. *Let $\mathcal{P}$ be the sequence of prime factors of n, counted with multiplicity. And let $\mathcal{T}$ be the subsequence of prime r in $\mathcal{P}$ with $\overline{r} \notin H$, then denote by t the length of $\mathcal{T}$. Considering the projection map: $\mathbb{Z} \to \mathbb{Z}/a\mathbb{Z}$, suppose the image of the sequence $\mathcal{P}$ contains H.*

(i) If $t \geq 2^{m-1}$, then $n \notin \mathcal{E}_a^*$.

(ii) If $t = 2^{m-1} - 1$, then $n \in \mathcal{E}_a^*$ if and only if every prime factor in $\mathcal{T}$ is congruent

to $g^{e'}$ *modulo* p^γ *for a fixed primitive root* $g \pmod{p^\gamma}$, *and for some* e' *such that* $e' \equiv \pm e \pmod{2^m}$ *with* e *being a fixed odd number.*

Proof. Recall that G and H are give by (4.2) and (4.3) respectively, for a fixed primitive root g modulo a. Denote $\mathcal{T} = \{r_j\}_1^t$. Let the sequence $\{e_j\}$ be such that $g^{e_j} \equiv r_j \pmod{a}$. By the assumption $\overline{r_j} \notin H$ we know $e_j \not\equiv 0 \pmod{2^m}$, for $1 \le j \le t$. Let $\mathcal{G} = \mathbb{Z}/2^m\mathbb{Z}$. Now $\{e_j\}$ can be viewed as a sequence of nonzero elements in $\mathcal{G}$. Clearly we see Lemma 4.5 gets into play here. More precisely, when $t \ge 2^{m-1}$, there exist $\delta_j \in \{-1, 0, 1\}$ such that

$$\sum_{j=1}^{t} \delta_j e_j \equiv 2^{m-1} \pmod{2^m}.$$

This is equivalent to

$$\sum_{j=1}^{t} \delta_j e_j \equiv b2^{m-1} \pmod{2^m d},$$

for some odd number b such that $1 \le b \le d$. Hence

$$\sum_{j=1}^{t} \delta_j e_j + (d-b)2^{m-1} \equiv 2^{m-1}d \pmod{2^m d}.$$

Translating this using multiplicative language, we know that

$$g^{\frac{d-b}{2}\cdot 2^m} \prod_{j=1}^{t} (g^{e_j})^{\delta_j} \equiv g^{2^{m-1}d} \equiv -1 \pmod{a}.$$

By assumption there exists $q \in \mathcal{P}$ such that $q \equiv g^{\frac{d-b}{2}\cdot 2^m} \pmod{a}$. On the other hand, $g^{e_j} \equiv r_j \pmod{a}$ and $q\prod_{j=1}^{t} r_j | n$. Hence there exist two coprime divisors u and v of n, such that $\frac{u}{v} \equiv -1 \pmod{a}$ namely $u + v \equiv 0 \pmod{a}$. By Lemma 4.3, we know $n \notin \mathcal{E}_a^*$. This proves part (i).

For part (ii), the necessity of the condition follows by exactly the same argument as above, keeping in mind that Lemma 4.5 still plays an important role. Now in order to prove the sufficiency, we just need to reverse the above argument and argue by contradiction. (Notice that the condition $\overline{\mathcal{P}}$ contains H is not needed in this direction.) □

Our next task naturally is to extend Lemma 4.6 to general modulus. We will see how one can carry the arguments here to the general case only with some mild difficulties in the next subsection.

4.2.3 The case for general a

Now we treat the general case $a = 2^{\gamma_0} p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k}$. Of course by Chinese remainder theorem we have the group isomorphism

$$(\mathbb{Z}/a\mathbb{Z})^* \simeq (\mathbb{Z}/2^{\gamma_0}\mathbb{Z})^* \times (\mathbb{Z}/p_1^{\gamma_1}\mathbb{Z})^* \times \cdots \times (\mathbb{Z}/p_k^{\gamma_k}\mathbb{Z})^*.$$

As before, we still denote this group by G. Here all the groups $(\mathbb{Z}/p^{\gamma}\mathbb{Z})^*$ are cyclic when p is an odd prime, but in general $(\mathbb{Z}/2^{\gamma_0}\mathbb{Z})^*$ is not except that $\gamma_0 \leq 2$. For instance, $(\mathbb{Z}/2\mathbb{Z})^*$ is trivial and $(\mathbb{Z}/4\mathbb{Z})^*$ is isomorphic to $(\mathbb{Z}/2\mathbb{Z}, +)$. In particular, there is no difference between the cases $\gamma_0 = 0$ and $\gamma_0 = 1$ because they exert no influence to G. While, when $\gamma_0 \geq 3$, $(\mathbb{Z}/2^{\gamma_0}\mathbb{Z})^*$ is a product of 2 cyclic groups with generators $-1 \pmod{2^{\gamma_0}}$ and $5 \pmod{2^{\gamma_0}}$ respectively, namely

$$(\mathbb{Z}/2^{\gamma_0}\mathbb{Z})^* \simeq \langle\overline{-1}\rangle \times \langle\overline{5}\rangle.$$

Apparently $|\langle\overline{-1}\rangle| = 2$ and $|\langle\overline{5}\rangle| = 2^{\gamma_0 - 2}$.

Here, we still want to find a maximal subgroup H of G such that $-1 \pmod{a} \notin H$. However, the issue here is that such subgroups of G might not be unique. They can be easily constructed as follows. Let $G_i = (\mathbb{Z}/p_i^{\gamma_i}\mathbb{Z})^*$, for $0 \leq i \leq k$ and H_i to be the maximal subgroup of G_i such that $-1 \pmod{p_i^{\gamma_i}} \notin H_i$. As we remarked before, $H_1, H_2, \cdots, H_k$ are unique but H_0 is not in general. In fact, H_0 is trivial if $\gamma_0 \leq 2$ and is one of the two subgroups of index 2 in the ambient group G_0 if $\gamma_0 \geq 3$. Recall our discussion in the cyclic case, hence $[G_i : H_i] = 2^{m_i}$ for some positive integer m_i and for all $1 \leq i \leq k$. Moreover

$$[G_0 : H_0] = \begin{cases} 1, & \text{if} \quad \gamma_0 \leq 1 \\ 2, & \text{if} \quad \gamma_0 \geq 2. \end{cases}$$

Now choose m such that

$$m = \begin{cases} \min\limits_{1\le i\le k} m_i, & \text{if} \quad \gamma_0 \le 1 \\ \\ 1, & \text{if} \quad \gamma_0 \ge 2, \end{cases}$$

namely

$$2^m || \gcd(\delta, p_1 - 1, p_2 - 1, \cdots, p_k - 1),$$

where

$$\delta = \begin{cases} 0, & \text{if} \quad \gamma_0 \le 1, \\ 2, & \text{if} \quad \gamma_0 \ge 2. \end{cases}$$

By definition, we have $m \ge 1$. The subgroup H as described above is one of the following groups with index $[G : H] = 2^m$:

$$H_0 \times G_1 \times \cdots \times G_k, G_0 \times H_1 \times \cdots \times G_k, \cdots$$

in which we just replace the i-th component of G by H_i for $0 \le i \le k$. We write $\phi(a) = 2^m d$ and hence $|H| = \frac{\phi(a)}{2^m} = d$. Notice that d is not necessarily odd in general.

It's routine to prove the following lemma (see the proof of Lemma 4.4).

Lemma 4.7. *We have the following inclusion relation of sets*

$$\{n \in \mathbb{N} : q|n \text{ with } q \text{ being prime } \Rightarrow \overline{q} \in H\} \subseteq \mathcal{E}_a^*.$$

The next lemma is crucial for our arguments.

Lemma 4.8. *Let H' be a subset of G with cardinality $|H'| \ge d$. And suppose for each $h' \in H'$, there are at least $\phi(a)$ many (counted with multiplicity) prime factors q of n satisfying $q \equiv h' \pmod a$. Then $n \notin \mathcal{E}_a^*$ unless $H' = H$ for some subgroup H defined above.*

Proof. The proof still relies on Lemma 4.3. Actually, by Lemma 4.3, if we can find a divisor of n which is congruent to $-1 \pmod a$, then $n \notin \mathcal{E}_a^*$. Now our argument goes roughly as follows, the fact that n has sufficiently many primes factors lying in sufficiently many different reduced residue classes in G, forces n to have at least one divisor lying in the residue class $-1 \pmod a$ unless H' is one of the above subgroups of G. To make this statement rigorous, let H'' be the set of all the residue classes of divisors of n in G. Then $\overline{1} \in H''$ and for any two elements $h_1, h_2 \in H'$, we have $h_1^{-1} = h_1^{\phi(a)-1} \in H''$ and $h_1 h_2 \in H''$. This means that H'' contains the subgroup $\langle H' \rangle$ generated by the elements in H' and in particular this subgroup has cardinality at least $|H'| \geq d$. However our H is maximized such that $\overline{-1} \notin H$, which implies either that $\overline{-1} \in \langle H' \rangle$ and hence $\overline{-1} \in H''$, or that H' itself is a maximal subgroup such that $\overline{-1} \notin H'$. In the former case, we have $n \notin \mathcal{E}_a^*$ by Lemma 4.3 and in the latter case, we know by Lemma 4.7 that $n \in \mathcal{E}_a^*$. □

Now we need an analogue of Lemma 4.6 for the general case. Here we need to pay special attention to the power of 2 dividing a. When $\gamma_0 \geq 2$, we have $m = 1$, which is sort of the "worst" case, for $E_a(N)$ is largest possible.

Lemma 4.9. *Suppose $H = G_0 \times G_1 \times \cdots \times H_i \times \cdots \times G_k$ is a subgroup of G defined as above. Let $\mathcal{P}$ be the sequence of prime factors of n (counted with multiplicity). And let $\mathcal{T}$ be the subsequence of prime r in $\mathcal{P}$ with $\overline{r} \notin H$. Then denote by t the length of $\mathcal{T}$. Considering the projection map: $\mathbb{Z} \to \mathbb{Z}/a\mathbb{Z}$, suppose the image of the sequence $\mathcal{P}$ contains H.*

(i) If $t \geq 2^{m-1}$, then $n \notin \mathcal{E}_a^*$.

(ii) If $t = 2^{m-1} - 1$ and $m \geq 2$ (in this case, $\gamma_0 \leq 1$ and hence G_0 is trivial and in particular our H_i here cannot be H_0), then $n \in \mathcal{E}_a^*$ if and only if every prime factor in $\mathcal{T}$ is congruent to $g^{e'}$ modulo $p_i^{\gamma_i}$ for a fixed primitive root $g \pmod{p_i^{\gamma_i}}$, and for some e' such that $e' \equiv \pm e \pmod{2^m}$ with e being a fixed odd number.

Proof. Generally speaking the arguments in the proof of Lemma 4.6 still work here. Nevertheless, one needs to make some changes accordingly. It is trivial to verify the conclusions when $m = 1$. So without loss of generality we assume $m \geq 2$ hence $1 \leq i \leq k$.

Denote $\mathcal{T} = \{r_j\}_1^t$ and fix a primitive root g modulo $p_i^{\gamma_i}$. Let the sequence $\{e_j\}$ be such that $g^{e_j} \equiv r_j \pmod{p_i^{\gamma_i}}$. By the assumption $r_j \pmod a \notin H$ namely $r_j \pmod{p_i^{\gamma_i}} \notin H_i$ we know $e_j \not\equiv 0 \pmod{2^m}$, for $1 \leq j \leq t$. Let $\mathcal{G} = \mathbb{Z}/2^m\mathbb{Z}$. Now

$\{e_j\}$ can be viewed as a sequence of nonzero elements in $\mathcal{G}$. Hence by Lemma 4.5, when $t \geq 2^{m-1}$, there exist $\delta_j \in \{-1, 0, 1\}$ such that

$$\sum_{j=1}^{t} \delta_j e_j \equiv 2^{m-1} \pmod{2^m}.$$

After writing $\phi(p_i^{\gamma_i}) = 2^m d_i$ with d_i odd. This is equivalent to

$$\sum_{j=1}^{t} \delta_j e_j \equiv b2^{m-1} \pmod{2^m d_i},$$

for some odd number b such that $1 \leq b \leq d_i$. Hence

$$\sum_{j=1}^{t} \delta_j e_j + (d_i - b)2^{m-1} \equiv 2^{m-1} d_i \pmod{2^m d_i}.$$

Translating this using multiplicative language, we know that

$$g^{\frac{d_i - b}{2} \cdot 2^m} \prod_{j=1}^{t} (g^{e_j})^{\delta_j} \equiv g^{2^{m-1} d_i} \equiv -1 \pmod{p_i^{\gamma_i}}.$$

By assumption there exists $q \in \mathcal{P}$ such that

$$\begin{cases} q \equiv -\prod_{j=1}^{t} r_j^{-\delta_j} \pmod{p_j^{\gamma_j}}, \quad 1 \leq j \leq k, j \neq i \\ \\ q \equiv g^{\frac{d_i - b}{2} \cdot 2^m} \pmod{p_i^{\gamma_i}}. \end{cases}$$

Hence by the Chinese remainder theorem, we know

$$q \prod_{j=1}^{t} r_j^{\delta_j} \equiv -1 \pmod{a}$$

namely there exist two coprime divisors u and v of n, such that $\frac{u}{v} \equiv -1 \pmod{a}$ namely $u + v \equiv 0 \pmod{a}$. Again by Lemma 4.3, we know $n \notin \mathcal{E}_a^*$.

Part (ii) can be proved similarly (see the comment in the proof of Lemma 4.6). □

4.3 The Analytic Inputs

We need the following generalisation of Ikehara's Tauberian Theorem, which is due to Delange ([3], see also Theorem 7.15 in Tenenbaum [35]). This extends Ikehara's Theorem to the case of a singularity of mixed type, involving algebraic and logarithmic poles. As usual we use σ to denote the real part of the complex number s, and we define $l(s) = \log \frac{1}{s-1}$ for $\sigma > 1$ by taking $l(2) = 0$ and then defining $l(s)$ by continuous variation along the line segment joining 2 to s.

Lemma 4.10 (Delange, 1954). *Let $f(s) = \sum_{n=1}^{\infty} a_n n^{-s}$ be a Dirichlet series with non-negative coefficients, converging for $\sigma > 1$. Suppose that $f(s)$ is holomorphic at all points of the line $\sigma = 1$ other than $s = 1$ and that, in the neighborhood of the this point and for $\sigma > 1$, we have*

$$f(s) = (s-1)^{-\omega-1} \sum_{j=0}^{t} g_j(s) \left(\log \left(\frac{1}{s-1} \right) \right)^j + g(s),$$

where ω is some real number, and the $g_j(s)$ and $g(s)$ are functions holomorphic at $s = 1$, the number $g_t(1)$ being non-zero. Then:

(i) if ω is not a negative integer, we have as $x \to \infty$

$$\sum_{n \leq x} a_n \sim \frac{g_t(1)}{\Gamma(\omega+1)} x (\log x)^{\omega} (\log \log x)^t,$$

(ii) if $\omega = -m-1$ for a non-negative integer m and if $t \geq 1$, we have as $x \to \infty$

$$\sum_{n \leq x} a_n \sim (-1)^m m! t g_t(1) x (\log x)^{\omega} (\log \log x)^{t-1}.$$

The following lemma is the key analytic ingredient of this paper. Essentially it plays the role of a sieve, but the upshot is that it produces asymptotics, not just an upper bound as almost all sieves do.

Lemma 4.11. *Suppose a is a positive integer, and let $\mathcal{B} = \{\overline{b_1}, \ldots, \overline{b_w}\}$ be a subset of $(\mathbb{Z}/a\mathbb{Z})^*$ with $w \geq 0$ elements, $\mathcal{C} = \{\overline{c_j}\}_1^t$ be a sequence of length t with elements in $(\mathbb{Z}/a\mathbb{Z})^*$ (elements could be repeated). And suppose further that $\mathcal{B}$ and $\mathcal{C}$ do not* *share common elements. Now let $\mathbb{P}$ denote the set of primes and define*

$$\mathcal{A} = \mathcal{A}(\mathcal{B}, \mathcal{C}) = \{q_1 q_2 \ldots q_l r_1 r_2 \ldots r_t : q_i \in \mathbb{P}, r_j \in \mathbb{P}, \overline{q_i} \in \mathcal{B}, \overline{r_j} = \overline{c_j}, l \geq 0\}.$$

Then

(i) if $w \geq 1$, we have as $x \to \infty$

$$|\mathcal{A}(x)| \sim C(a, \mathcal{B}, t) \frac{x(\log\log x)^t}{(\log x)^{1-w/\phi(a)}},$$

(ii) if $w = 0$ and $t \geq 1$, we have as $x \to \infty$

$$|\mathcal{A}(x)| \sim C(a, t) \frac{x(\log\log x)^{t-1}}{\log x}.$$

The constants $C(a, \mathcal{B}, t)$ and $C(a, t)$ are positive and do not depend on the choices of the $\overline{c_j}$.

Proof. Let
$$a_n = \begin{cases} 1, & n \in \mathcal{A}, \\ 0, & n \notin \mathcal{A}. \end{cases}$$

The set $\mathcal{A}(\mathcal{B}, \mathcal{C})$ has a multiplicative structure, and this leads naturally to the following Dirichlet series

$$f(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s} = \prod_{\substack{q \in \mathbb{P} \\ \overline{q} \in \mathcal{B}}} (1 - 1/q^s)^{-1} \prod_{j=1}^{t} \sum_{\substack{r \in \mathbb{P} \\ \overline{r} = \overline{c_j}}} \frac{1}{r^s} \tag{4.4}$$

which converges absolutely and locally uniformly in the region $\sigma > 1$.

When $D(s)$ is a Dirichlet series which converges absolutely and locally uniformly for $\sigma > \sigma_0$, has an analytic continuation for $\sigma > \sigma_1$, is non-zero for $\sigma > \sigma_2$ and satisfies $\lim_{\sigma \to \infty} D(\sigma) = 1$, we define $D(s)^\alpha$ for $\sigma > \max(\sigma_1, \sigma_2)$ and an arbitrary complex number α by $\exp(\alpha \log D(s))$ where we choose the principal value of $\log D(\sigma_3)$ for some suitably large σ_3 and then define $\log D(s)$ by continuous variation from σ_3 to s.

Let $e(\chi) = \frac{1}{\phi(a)} \sum_{\chi \bmod a} \chi(q)\bar{\chi}(b)$. Then, by the orthogonality of Dirichlet characters, the product over q on the right of the equation (4.4) is

$$\begin{aligned}&\prod_{\bar{b}\in\mathcal{B}}\prod_{q}(1-1/q^s)^{-e(\chi)}\\=&\prod_{\bar{b}\in\mathcal{B}}\prod_{\chi \bmod a}(L(s,\chi)g_1(s,\chi))^{\frac{\bar{\chi}(b)}{\phi(a)}}\end{aligned}$$

where

$$g_1(s,\chi)=\prod_{q}\frac{(1-\chi(q)/q^s)}{(1-1/q^s)^{\chi(q)}},$$

which converges absolutely when $\sigma>\frac{1}{2}$, and hence has no zeros in that region. Thus $g_1(s,\chi)^{\frac{\bar{\chi}(b)}{\phi(a)}}$ is a well defined analytic function when $\sigma>\frac{1}{2}$.

Now the above product can be further rearranged as

$$L(s,\chi_0)^{\frac{\omega}{\phi(a)}}g_1(s) \tag{4.5}$$

where ω is the cardinality of $\mathcal{B}$, χ_0 is the principal character modulo a and

$$g_1(s)=\prod_{\bar{b}\in\mathcal{B}}\prod_{\substack{\chi\neq\chi_0\\ \bmod a}}(L(s,\chi)g_1(s,\chi))^{\frac{\bar{\chi}(b)}{\phi(a)}}.$$

In particular

$$g_1(1)\neq 0.$$

Note that $g_1(1)$ may depend on the choice of $\mathcal{B}$.

It is well known that $L(s,\chi)$ has no zeros with $\sigma\geq 1$ and has an analytic continuation to the whole complex plane. Moreover, when χ is non-principal it is entire and when χ is a principal character χ_0 it has a simple pole at $s=1$ and $(s-1)L(s,\chi_0)$ is entire. Thus, when χ is non-principal ,

$$L(s,\chi)^{\frac{\bar{\chi}(b)}{\phi(a)}}$$

is analytic in the region $\sigma\geq 1$ and hence so is $g_1(s)$.

On the other hand, again by the orthogonality of Dirichlet characters the sum over r on the right of the equation (4.4) is

$$\frac{1}{\phi(a)} \sum_{\chi \bmod a} \bar{\chi}(-c_j) \sum_{p} \frac{\chi(p)}{p^s}.$$

Now it is readily verified that when $\sigma > 1$ we have

$$\begin{aligned} \log L(s,\chi) &= -\sum_{p} \log\left(1 - \frac{\chi(p)}{p^s}\right) \\ &= \sum_{p} \frac{\chi(p)}{p^s} + \sum_{p} \sum_{k=2}^{\infty} \frac{\chi(p^k)}{p^{ks}}. \end{aligned}$$

The second sum on the right converges locally uniformly when $\sigma > \frac{1}{2}$. Thus

$$\sum_{p} \frac{\chi(p)}{p^s} = \log L(s,\chi) + h(s,\chi)$$

where $h(s,\chi)$ is holomorphic for $\sigma > \frac{1}{2}$. Notice that $\log L(s,\chi)$ is analytic on the line $\sigma = 1$ except when $\chi = \chi_0$ when it has a logarithmic singularity at the point $s = 1$. Hence

$$\sum_{\substack{p \\ \overline{p} = \overline{c_j}}} \frac{1}{p^s} = \frac{1}{\phi(a)} \log L(s,\chi_0) + h(s,c_j)$$

where $h(s,c_j)$ is an analytic function of s for $\sigma \geq 1$. Therefore

$$\prod_{j=1}^{t} \sum_{\substack{p \\ \overline{p} = \overline{c_j}}} \frac{1}{p^s} = \frac{1}{\phi(a)^t} \sum_{j=0}^{t} (\log L(s,\chi_0))^j h_j(s) \tag{4.6}$$

where the $h_j(s)$ are analytic when $\sigma \geq 1$ and $h_t(1) = 1$.

Now on combining (4.4), (4.5) and (4.6), we have

$$f(s) = \frac{g_1(s)}{\phi(a)^t} L(s,\chi_0)^{\frac{w}{\phi(a)}} \sum_{j=0}^{t} (\log L(s,\chi_0))^j h_j(s).$$

We have $L(s,\chi_0) = \zeta(s) \prod_{p|a} (1 - p^{-s})$ and the Riemann zeta function $\zeta(s)$ has a simple pole at $s = 1$ with residue 1 at $s = 1$. Thus

$$L(s, \chi_0) = \frac{\phi(a) g_2(s)}{a(s-1)},$$

where $g_2(s)$ is an entire function with $g_2(1) = 1$. On plugging this in to the above expression for $f(s)$, the asymptotic formula of Lemma 4.11 follows from Lemma 4.10. Notice that we apply part (i) of Lemma 4.10 when $w \geq 1$ and part (ii) when $w = 0$ and $t \geq 1$. That the constants $C(a, \mathcal{B}, t)$ and $C(a, t)$ are positive follows by observing first that, by Lemma 4.10, they are non-zero and then that the left hand side of the asymptotic formula is non-negative. □

4.4 Proof of Theorem 4.1

The main analytic tool in the proof of Theorem 4.1 is Lemma 4.11 and we will apply it to the various sets from Section 4.2.3. Recall the definitions of the groups G and H and of the numbers m and d from Section 4.2.3. We denote by $\mathfrak{H}$ the set of all subgroups H defined in Section 4.2.3 for general a. Now as was defined in Lemma 4.11, we form the set

$$\mathcal{A}(H, \mathcal{C}),$$

where

$$H = G_0 \times G_1 \times \cdots \times H_i \times \cdots \times G_k \in \mathfrak{H}$$

and $\mathcal{C} = \{\overline{c_j}\}_1^t$ is a sequence of length $t = 2^{m-1} - 1$ with elements in G. Moreover for a fixed primitive root $g \pmod{p_i^{\gamma_i}}$ and a fixed odd number e we have $c_j \equiv g^{e'} \pmod{p_i^{\gamma_i}}$ for some e' with $e' \equiv \pm e \pmod{2^m}$. For a fixed $H \in \mathfrak{H}$, there are only finitely many possibilities ($d^t 2^{t+m-1}$ actually) for $\mathcal{C}$. Lemma 4.11 immediately implies that

Lemma 4.12.

$$|\mathcal{A}(H, \mathcal{C})(N)| \sim C(H, \mathcal{C}) \frac{N (\log\log N)^{2^{m-1}-1}}{(\log N)^{1-1/2^m}}.$$

Now we need to show that the intersection of any two distinct such sets, $\mathcal{A}(H^1, \mathcal{C}^1)$ and $\mathcal{A}(H^2, \mathcal{C}^2)$, is a relatively small set.

Lemma 4.13. *We have*

$$|\left(\mathcal{A}(H^1,\mathcal{C}^1)\cap\mathcal{A}(H^2,\mathcal{C}^2)\right)(N)| \ll_a \frac{N(\log\log N)^{2^m-2}}{(\log N)^{1-1/4^m}}$$

Proof. If H^1 and H^2 are the same, then $\mathcal{C}^1$ and $\mathcal{C}^2$ differs in at least one element. Hence the intersection is empty. So without loss of generality, we can assume H^1 and H^2 are not the same. Then $H^1\cap H^2$ is a subgroup of G with index 4^m. Also notice the relation

$$\mathcal{A}(H^1,\mathcal{C}^1)\cap\mathcal{A}(H^2,\mathcal{C}^2)\subseteq\mathcal{A}(H^1\cap H^2,\mathcal{C}^1\cup\mathcal{C}^2)$$

where $\mathcal{C}^1\cup\mathcal{C}^2$ is the union of the sequences $\mathcal{C}^1$ and $\mathcal{C}^2$ and hence is of length 2^m-2. Then the desired conclusion follows from Lemma 4.11. □

Now set

$$\mathcal{U}=\cup_H\cup_{\mathcal{C}}\mathcal{A}(H,\mathcal{C}),$$

where the union runs through all $H\in\mathfrak{H}$ and the corresponding sequences $\mathcal{C}$ for H as defined above. We know that $\mathcal{U}\subseteq\mathcal{E}_a^*$ from Lemma 4.9.

Lemma 4.14. *We have*

$$E_a^*(N)-|\mathcal{U}(N)|\ll_a\begin{cases}\frac{N(\log\log N)^{2^{m-1}-2}}{(\log N)^{1-1/2^m}}, & \text{if } m\geq 2,\\ \frac{N(\log\log N)^{\phi(a)^{\phi(a)}}}{(\log N)^{1-1/2^m+1/\phi(a)}}, & \text{if } m=1.\end{cases}$$

Proof. We let $\mathcal{W}(n)$ be the set of residue classes modulo a in which there are at least $\phi(a)$ (counted with multiplicity) prime factors of n. By Lemma 4.8 we know that

(i) if $|\mathcal{W}(n)|\geq d+1$, then $n\notin\mathcal{E}_a^*$;

(ii) if $|\mathcal{W}(n)|=d$, then $n\notin\mathcal{E}_a^*$ unless $\mathcal{W}(n)=H$ for some subgroup H of G as above.

Let

$$\mathcal{N}(i)=\{n\in\mathcal{E}_a^*:|\mathcal{W}(n)|=i\}$$

for $0 \leq i \leq \phi(a)$. From the above discussion we know $\mathcal{N}(i)$ is empty as long as $i > d$. Hence

$$\mathcal{E}_a^* = \bigcup_{i=0}^{d} \mathcal{N}(i).$$

Firstly observe that by Lemma 4.11 we have

$$\left| \left(\bigcup_{i=0}^{d-1} \mathcal{N}(i) \right) (N) \right| \ll_a \frac{N(\log\log N)^{\phi(a)^{\phi(a)}}}{(\log N)^{1-1/2^m+1/\phi(a)}}.$$

Now if $m = 1$, then we have $\mathcal{N}(d) = \mathcal{U}$ by part (i) of Lemma 4.9, and if $m \geq 2$, then we have by Lemma 4.9 and Lemma 4.11 that

$$|(\mathcal{N}(d))(N)| - |\mathcal{U}(N)| \ll_a \frac{N(\log\log N)^{2^{m-1}-2}}{(\log N)^{1-1/2^m}}.$$

Therefore Lemma 4.14 follows by putting the above conclusions together. □

Here we bound the error term rather crudely, following from Lemma 4.8. Actually it can be refined substantially by a generalisation of Lemma 4.5, which is, however, not pertinent to the purpose of the current paper.

Now Theorem 4.1 follows from Lemma 4.2, Lemma 4.12, Lemma 4.13 and Lemma 4.14. It should be noted that the leading constant $C(a)$ appearing in Theorem 4.1, can be traced back explicitly in our arguments, but is inevitably messy, would require some non-trivial expenditure of effort and would not give any further insights into our problem.

Chapter 5

A Mean Value Theorem for the Diophantine Equation $axy - x - y = n$

5.1 Introduction

People has been considering Diophantine equations involving products and sums of some variables for a long time. The Diophantine equation

$$\prod_{i=1}^{k} x_i - \sum_{i=1}^{k} x_i = n \tag{5.1}$$

was studied by various people during the past a few decades. It is easy to see that there always exists a few trivial solutions with most of x_i's equal to 1. So people are asking about the number of solutions of this equation with all $x_i > 1$.

The case when $n = 0$ is very special, since it concerns the number of k-tuples with equal sum and product. In this case, it is conjectured by Misiurewicz [26] that $k = 2, 3, 4, 6, 24, 114, 174$ and 444 are the only values of k for which there are only trivial solutions. For general n, very little is known except that in 1970s Viola [39] proved that if $E_k(N)$ denotes the number of positive integers $n \le N$ for which (5.1) is not soluble in integers $x_1, x_2, \ldots, x_k > 1$ then $E_k(N) = N \exp(-c_k (\log N)^{1-1/(k+1)})$ for some positive constant c_k. It is believed that for large n equation (5.1) always has a nontrivial solution, which nevertheless is an open question in this area.

On the other hand, the case that $k = 3$ has received extensive attention, and several variations of this problem were studied. Brian Conrey asked whether the number of solutions in positive integers to the equation

$$xyz + x + y = n \tag{5.2}$$

can be bounded by $O_\varepsilon(n^\varepsilon)$ for any $\varepsilon > 0$. Kevin Ford posed a generalisation of this problem, in which one would like to show that there are $O_\varepsilon(|AB|^\varepsilon)$ nontrivial positive integer solutions to the equation

$$xyz = A(x + y) + B \tag{5.3}$$

for given nonzero $A, B \in \mathbb{Z}$.

In this chapter, we consider another variation of the case that $k = 2$, namely the following equation

$$axy - x - y = n \tag{5.4}$$

where a is a positive integer and n is any nonnegative integer. This can be viewed as equation (5.3) in which z is fixed and $A = 1$. Hence if the number of solutions of equation (5.4) is well understood, then one can probably understand the number of solutions of equation (5.3) simply by averaging over a.

Let

$$R_a(n) = \mathrm{Card}\left\{(\mathrm{x}, \mathrm{y}) \in \mathbb{N}^2 : \mathrm{axy} - \mathrm{x} - \mathrm{y} = \mathrm{n}\right\}.$$

Here we are considering the number of positive integer solutions of equation (5.4) when a is fixed and n varies. A sharp asymptotic formula is established in this chapter on the average of $R_a(n)$ over n. Notice the case that $a = 1$ is trivial, since then $R_1(n) = d(n+1)$ is just the divisor function of $n+1$, the average of which is relatively well understood.

Theorem 5.1. *For positive integers $a > 1$ and $N \geq 1$, we have*

$$\sum_{0 \leq n \leq N} R_a(n) = \frac{1}{a}\Big(N \log N - C(a)N\Big) + \Delta_a(N)$$

where

$$C(a) = 2\frac{\Gamma'(\frac{a-1}{a})}{\Gamma(\frac{a-1}{a})} + 2\sum_{p|a}\frac{\log p}{p-1} + \log a + 2\gamma + 1 \tag{5.5}$$

and

$$\Delta_a(N) \ll \phi(a)\sqrt{\frac{N}{a}}\Big(\log(aN)\Big)^2. \tag{5.6}$$

Here $\Gamma(s) = \int_0^{+\infty} e^{-t}t^{s-1}dt$ *is the standard* Γ *function, and* γ *is the Euler constant.*

In fact, since the error term above is roughly of size $\sqrt{aN}\Big(\log(aN)\Big)^2$, it is conceivable that the main term will be inferior to the error term when $a \gg N^{\frac{1}{3}}$. So in order for the above asymptotic formula to really make sense, one would impose a condition on a, such as $a \ll N^{\frac{1}{3}}/\log N$.

Moreover, one can argue what is the right order of magnitude of the error $\Delta_a(N)$. In view of $R_1(n) = d(n+1)$, one can think $R_a(n)$ as a "generalized" divisor function. Hence Theorem 5.1 just proves a mean value theorem for such a "generalized" divisor function. Since for the classical divisor function, the error is believed to be $O(N^{1/4+\varepsilon})$. It is very natural to pose such a conjecture for our error $\Delta_a(N)$. The author suspects that following the van der Corput method on exponential sums as in the classical case, one can show $\Delta_a(N) = O_a(N^{1/3-\delta})$ for some $\delta > 0$.

Remark 5.1. It's not hard to adapt the method in this chapter in order to deal with equations like

$$axy - bx - cy = n$$

and prove similar asymptotic formulas.

5.2 Preliminary Lemmas

We state several lemmas before embarking on the proof of Theorem 1.

Lemma 5.2. *Let* $T \geq 2$*, then we have*

$$\sum_{\substack{\chi \\ \bmod a}} \int_{-T}^{T} |L(\frac{1}{2}+it,\chi)|^2 dt \ll \frac{\phi^2(a)}{a} T\log T.$$

A proof of this lemma can be found for example in Montgomery [27].

Lemma 5.3. *Let a be a positive integer greater than 1 and $w > 0$, we have*

$$\sum_{\substack{n \le w \\ n \equiv -1 \bmod a}} \frac{1}{n} = \frac{1}{a}\left(\log w - \frac{\Gamma'(\frac{a-1}{a})}{\Gamma(\frac{a-1}{a})} - \log a\right) + O(1/w).$$

Proof. By Abel summation, the left hand side above is

$$\begin{aligned} \sum_{\substack{n \le w \\ n \equiv -1 \bmod a}} \frac{1}{n} &= \left\lfloor \frac{w+1}{a} \right\rfloor \frac{1}{w} + \int_1^w \left\lfloor \frac{t+1}{a} \right\rfloor \frac{1}{t^2} dt \\ &= \frac{1}{a} + \int_1^w \frac{t+1}{at^2} dt - \int_1^w \left\{ \frac{t+1}{a} \right\} \frac{dt}{t^2} + O(1/w) \\ &= \frac{1}{a}\left(\log w + 2 - \int_1^\infty a \left\{ \frac{t+1}{a} \right\} \frac{dt}{t^2}\right) + O(1/w). \end{aligned}$$

Recall that the digamma function $\psi(z)$ is defined as $\frac{\Gamma'}{\Gamma}(z)$, and $\psi'(z)$ has a series expansion $\sum_{k=0}^\infty \frac{1}{(z+k)^2}$. So

$$\begin{aligned} \int_1^\infty \left(a\left\{\frac{t+1}{a}\right\} - \{t\} - 1\right) \frac{dt}{t^2} &= \sum_{h=0}^\infty \int_0^a \left(a\left\{\frac{r+1}{a}\right\} - \{r\} - 1\right) \frac{dr}{(ah+r)^2} \\ &= \frac{1}{a^2} \int_0^a \left(a\left\{\frac{r+1}{a}\right\} - \{r\} - 1\right) \psi'\left(\frac{r}{a}\right) dr \end{aligned} \tag{5.7}$$

Notice that

$$a\left\{\frac{r+1}{a}\right\} - \{r\} - 1 = \begin{cases} 0, & \text{if} \quad 0 \le r < 1 \\ 1, & \text{if} \quad 1 \le r < 2 \\ \vdots & \qquad \vdots \\ a-2, & \text{if} \quad a-2 \le r < a-1 \\ -1, & \text{if} \quad a-1 \le r < a \end{cases}$$

Hence (5.7) is equal to

$$\begin{aligned}
&\frac{1}{a^2}\left(\sum_{l=1}^{a-2} l \int_0^1 \psi'\left(\frac{l+r}{a}\right)dr - \int_0^1 \psi'\left(\frac{a-1+r}{a}\right)dr\right)\\
= \ &\frac{1}{a}\left(\sum_{l=1}^{a-2} l \left(\psi\left(\frac{l+1}{a}\right) - \psi\left(\frac{l}{a}\right)\right) - \left(\psi(1) - \psi\left(\frac{a-1}{a}\right)\right)\right)\\
= \ &\psi\left(\frac{a-1}{a}\right) - \frac{1}{a}\sum_{l=1}^{a}\psi\left(\frac{l}{a}\right)\\
= \ &\psi\left(\frac{a-1}{a}\right) + \log a + \gamma
\end{aligned}$$

The last equality follows from a well known property of the digamma function ψ. Now the lemma is established after the observation $\gamma = 2 - \int_1^\infty \frac{\{t\}+1}{t^2}dt$. □

Lemma 5.4. *Let a be a positive integer greater than 1, then we have*

$$\frac{1}{\phi(a)} \sum_{\substack{\chi \neq \chi_0 \\ \text{mod } a}} \bar{\chi}(-1) L(1,\chi) = -\frac{1}{a}\left(\frac{\Gamma'(\frac{a-1}{a})}{\Gamma(\frac{a-1}{a})} + \sum_{p|a}\frac{\log p}{p-1} + \log a + \gamma\right).$$

Proof. Let w be large compared to a (eventually we will let w goes to ∞). Then for non-principal characters χ modulo a, by Abel summation

$$L(1,\chi) = \sum_{n \leq w}\frac{\chi(n)}{n} + O(a/w).$$

Hence

$$\begin{aligned}
&\frac{1}{\phi(a)} \sum_{\substack{\chi \neq \chi_0 \\ \text{mod } a}} \bar{\chi}(-1) L(1,\chi)\\
&\qquad = \frac{1}{\phi(a)} \sum_{\substack{\chi \neq \chi_0 \\ \text{mod } a}} \bar{\chi}(-1) \sum_{n \leq w}\frac{\chi(n)}{n} + O(a/w).
\end{aligned}$$

The main term on the right is

$$\frac{1}{\phi(a)} \sum_{\substack{\chi \\ \bmod a}} \bar{\chi}(-1) \sum_{n \le w} \frac{\chi(n)}{n} - \frac{1}{\phi(a)} \sum_{\substack{n \le w \\ (n,a)=1}} \frac{1}{n}.$$

We have

$$\begin{aligned}
\sum_{\substack{n \le w \\ (n,a)=1}} \frac{1}{n} &= \sum_{m|a} \frac{\mu(m)}{m} \sum_{n \le w/m} \frac{1}{n} \\
&= \sum_{m|a} \frac{\mu(m)}{m} \Big(\log(w/m) + \gamma + O(m/w) \Big) \\
&= \frac{\phi(a)}{a} \left(\log w + \sum_{p|a} \frac{\log p}{p-1} + \gamma \right) + O\Big(d(a)/w\Big).
\end{aligned}$$

Here we are using the fact that $-\sum_{m|a} \frac{\mu(m)}{m} \log m = \frac{\phi(a)}{a} \sum_{p|a} \frac{\log p}{p-1}$, this is because

$$\begin{aligned}
-\sum_{m|a} \frac{\mu(m)}{m} \log m &= \sum_{p|a} \frac{\log p}{p} \sum_{\substack{k|a/p \\ (p,k)=1}} \frac{\mu(k)}{k} \\
&= \sum_{p|a} \frac{\log p}{p} \prod_{\substack{p'|a \\ p' \ne p}} \left(1 - \frac{1}{p'}\right) \\
&= \sum_{p|a} \frac{\log p}{p} \left(\frac{1}{1 - \frac{1}{p}} \right) \prod_{p'|a} \left(1 - \frac{1}{p'}\right) \\
&= \frac{\phi(a)}{a} \sum_{p|a} \frac{\log p}{p-1}.
\end{aligned}$$

On the other hand, we have

$$\frac{1}{\phi(a)} \sum_{\substack{\chi \\ \bmod a}} \bar{\chi}(-1) \sum_{n \le w} \frac{\chi(n)}{n} = \sum_{\substack{n \le w \\ n \equiv -1 \bmod a}} \frac{1}{n}$$

And by lemma 5.3, this is

$$\frac{1}{a}\left(\log w - \frac{\Gamma'(\frac{a-1}{a})}{\Gamma(\frac{a-1}{a})} - \log a\right) + O(1/w).$$

Thus we have shown that

$$\begin{aligned} &\frac{1}{\phi(a)} \sum_{\substack{\chi\neq\chi_0 \\ \text{mod } a}} \bar{\chi}(-1)L(1,\chi) \\ = \;& \frac{1}{a}\left(\log w - \frac{\Gamma'(\frac{a-1}{a})}{\Gamma(\frac{a-1}{a})} - \log a\right) - \frac{1}{a}\left(\log w + \sum_{p|a}\frac{\log p}{p-1} + \gamma\right) + O(a/w) \\ = \;& -\frac{1}{a}\left(\frac{\Gamma'(\frac{a-1}{a})}{\Gamma(\frac{a-1}{a})} + \sum_{p|a}\frac{\log p}{p-1} + \log a + \gamma\right) + O(a/w) \end{aligned}$$

Now the lemma is established when we let $w \to \infty$ in the above. □

5.3 Proof of Theorem 5.1

The starting point of the proof is the following observation. One can rewrite equation (5.4) in the following form

$$(ax-1)(ay-1) = an+1. \tag{5.8}$$

Namely we are going to count the following quantities,

$$R_a(n) = \text{Card}\left\{(x,y) \in \mathbb{N}^2 : (ax-1)(ay-1) = an+1\right\}$$

and

$$S_a(N) = \sum_{0\leq n\leq N} R_a(n).$$

After the change of variables $u = ax-1$ and $v = ay-1$, it follows that $R_a(n)$ is the number of ordered pairs of natural numbers u, v such that $uv = an+1$ and $u \equiv v \equiv -1 \pmod a$.

Now the residue class $u \equiv -1 \pmod a$ and $v \equiv -1 \pmod a$ are readily isolated *via* the orthogonality of the Dirichlet characters χ modulo a. Thus we have

$$\begin{aligned} & S_a(N) \\ = & \sum_{0\le n\le N} \sum_{\substack{uv=an+1 \\ u\equiv -1 \bmod a \\ v\equiv -1 \bmod a}} 1 \\ = & \sum_{m\le M} \sum_{\substack{uv=m \\ u\equiv -1 \bmod a \\ v\equiv -1 \bmod a}} 1 \\ = & \frac{1}{\phi^2(a)} \sum_{\substack{\chi_1 \\ \bmod a}} \sum_{\substack{\chi_2 \\ \bmod a}} \bar{\chi}_1(-1)\bar{\chi}_2(-1) \sum_{m\le M} \sum_{uv=m} \chi_1(u)\chi_2(v), \end{aligned}$$

where $M = aN + 1$.

Let

$$a_m(\chi_1, \chi_2) = \sum_{uv=m} \chi_1(u)\chi_2(v).$$

Then we have

$$S_a(N) = \frac{1}{\phi^2(a)} \sum_{\substack{\chi_1 \\ \bmod a}} \sum_{\substack{\chi_2 \\ \bmod a}} \bar{\chi}_1(-1)\bar{\chi}_2(-1) \sum_{m\le M} a_m(\chi_1, \chi_2).$$

We analyze this expression through the properties of the Dirichlet series

$$f_{\chi_1,\chi_2}(s) = \sum_{m=1}^{\infty} \frac{a_m(\chi_1, \chi_2)}{n^s} = L(s, \chi_1)L(s, \chi_2). \tag{5.9}$$

This affords an analytic continuation of f_{χ_1,χ_2} to the whole complex plane.

By a quantitative version of Perron's formula, as in Theorem 5.2 of MV [28] for example, we obtain

$$\sum_{m\le M}{}' a_m(\chi_1, \chi_2) = \frac{1}{2\pi i} \int_{\sigma_0 - iT}^{\sigma_0 + iT} f_{\chi_1,\chi_2}(s) \frac{M^s}{s} ds + R(\chi_1, \chi_2),$$

where $\sigma_0 > 1$ and

$$\begin{aligned} R(\chi_1, \chi_2) \ll \ & \textstyle\sum_{\substack{\frac{M}{2}<m<2M \\ m\ne M}} |a_m(\chi_1, \chi_2)| \min\left(1, \frac{M}{T|m-M|}\right) \\ & + \tfrac{4^{\sigma_0}+M^{\sigma_0}}{T} \textstyle\sum_{m=1}^{\infty} \frac{|a_m(\chi_1,\chi_2)|}{m^{\sigma_0}}. \end{aligned}$$

Here $\sum'$ means that when M is an integer, the term $a_M(\chi_1,\chi_2)$ is counted with weight $\frac{1}{2}$.

Let $\sigma_0 = 1 + \frac{1}{\log M}$. By (5.9) we have $|a_m(\chi_1,\chi_2)| \le d(n)$. Thus

$$\sum_{m=1}^{\infty} \frac{|a_m(\chi_1,\chi_2)|}{n^{\sigma_0}} \ll \zeta(\sigma_0)^2 \ll (\log B)^2$$

and so $R(\chi_1,\chi_2) \ll_\varepsilon M^{1+\varepsilon}T^{-1}$, for any $\varepsilon > 0$. Hence

$$\sum_{m\le M} a_m(\chi_1,\chi_2) = \frac{1}{2\pi i}\int_{\sigma_0-iT}^{\sigma_0+iT} f_{\chi_1,\chi_2}(s)\frac{M^s}{s}ds + O\left(\left(\frac{M}{T}+1\right)M^\varepsilon\right).$$

The error term here is

$$\ll M^\varepsilon$$

provided that

$$T \ge M.$$

The integrand is a meromorphic function in the complex plane and is analytic for all s with $\Re s \ge \frac{1}{2}$ except for a possible pole of finite order at $s = 1$. Suppose that $T \ge 4$. By the residue theorem we have

$$\begin{aligned}
&\frac{1}{2\pi i}\int_{\sigma_0-iT}^{\sigma_0+iT} f_{\chi_1,\chi_2}(s)\frac{M^s}{s}ds \\
=&\frac{1}{2\pi i}\left(\int_{\sigma_0-iT}^{\frac{1}{2}-iT} + \int_{\frac{1}{2}-iT}^{\frac{1}{2}+iT} + \int_{\frac{1}{2}+iT}^{\sigma_0+iT}\right)\frac{L(s,\chi_1)L(s,\chi_2)M^s}{s}ds \\
&+ \mathrm{Res}_{s=1}\left(L(s,\chi_1)L(s,\chi_2)\frac{M^s}{s}\right).
\end{aligned}$$

Hence, by Lemmas 2.3, Lemma 2.4 and Lemma 2.5, the contribution from the horizontal paths is

$$\begin{aligned}
&\ll (\log aT)^2\frac{M}{T\log M} + \frac{(aT)^\varepsilon}{T}\int_{1/2}^{1}(aT)^{1-\sigma}M^\sigma d\sigma \\
&\ll T^{-1}(aT)^\varepsilon M + T^{-1}(aT)^{1/2+\varepsilon}M^{1/2}
\end{aligned}$$

and provided that $T \geq M^5$ this is

$$\ll M^{-1}.$$

On the other hand, the contribution from the vertical path on the right is bounded by

$$M^{\frac{1}{2}} \sum_{2^k \leq T} 2^{-k} \int_{2^k}^{2^{k+1}} |L(\tfrac{1}{2}+it,\chi_1)L(\tfrac{1}{2}+it,\chi_2)|dt.$$

And by Lemma 5.2

$$\sum_{\substack{\chi_1,\chi_2 \\ \bmod a}} \bar{\chi}_1(-1)\bar{\chi}_2(-1)\frac{1}{2\pi i}\int_{\frac{1}{2}-iT}^{\frac{1}{2}+iT} \frac{L(s,\chi_1)L(s,\chi_2)M^s}{s}ds$$

$$\ll M^{\frac{1}{2}} \sum_{2^k \leq T} 2^{-k} \int_{2^k}^{2^{k+1}} \Bigg(\sum_{\substack{\chi \\ \bmod a}} |L(\tfrac{1}{2}+it,\chi)| \Bigg)^2 dt$$

$$\ll M^{\frac{1}{2}} \sum_{2^k \leq T} 2^{-k}\phi(a) \sum_{\substack{\chi \\ \bmod a}} \int_{-2^{k+1}}^{2^{k+1}} |L(\tfrac{1}{2}+it,\chi)|^2 dt$$

$$\ll M^{\frac{1}{2}} \sum_{2^k \leq T} \frac{\phi^3(a)}{a}k$$

$$\ll \frac{\phi^3(a)}{a}M^{\frac{1}{2}}(\log M)^2$$

on taking

$$T = M^5.$$

Hence we obtain

$$S_a(N) = \frac{1}{\phi^2(a)} \sum_{\substack{\chi_1 \\ \bmod a}} \sum_{\substack{\chi_2 \\ \bmod a}} \bar{\chi}_1(-1)\bar{\chi}_2(-1)\mathrm{Res}_{s=1}\left(f_{\chi_1,\chi_2}(s)\frac{M^s}{s}\right) + \Delta_a(N)$$

where

$$\Delta_a(N) \ll \frac{\phi(a)}{a}\sqrt{M}(\log M)^2 \ll \phi(a)\sqrt{\frac{N}{a}}\Big(\log(aN)\Big)^2. \tag{5.10}$$

It remains to compute the residue at $s = 1$.

By (5.9) there are naturally two cases, namely

(i) $\chi_1 = \chi_2 = \chi_0$;

(ii) only one of χ_1 and χ_2 is equal to χ_0 while the other one is equal to $\chi \neq \chi_0$.

In the latter case the integrand has a simple pole at $s = 1$ and the residue is

$$\prod_{p|a}\left(1-\frac{1}{p}\right) L(1,\chi)(aN+1) = \phi(a)L(1,\chi)N + \frac{\phi(a)}{a}L(1,\chi).$$

By lemma 5.4, the sum over χ for the second term above is small, hence can be absorbed in $\Delta_a(N)$. While in the former case, the integrand has a double pole at $s = 1$ and the residue is

$$\prod_{p|a}\left(1-\frac{1}{p}\right)^2 \Big(M\log M - M\Big).$$

Hence we have shown that

$$\begin{aligned} S_a(N) = &\quad \tfrac{1}{a^2}\Big((aN+1)\log(aN+1) - aN - 1\Big) \\ + &\quad \Big(\frac{2}{\phi(a)} \sum_{\substack{\chi\neq\chi_0 \\ \bmod a}} \bar{\chi}(-1)L(1,\chi)\Big)N + \Delta_a(N). \end{aligned}$$

Now by lemma 5.4, this is

$$\frac{1}{a}\Big(N\log N - C(a)N\Big) + \Delta_a(N)$$

where $C(a)$ and $\Delta_a(N)$ are given by (5.5) and (5.6) respectively.

This completes the proof of Theorem 5.1.

□

Appendix A

The equivalence of Erdős' and Straus' conjectures

Proposition A.1. *When $n > 2$, then Equation* (1.2)

$$\frac{a}{n} = \frac{1}{x} + \frac{1}{y} + \frac{1}{z}$$

has positive integer solutions if and only if it has pairwise distinct positive integer solutions.

Proof. We apparently have

$$\frac{4}{3} = 1 + \frac{1}{4} + \frac{1}{12}$$

and

$$\frac{4}{4} = \frac{1}{2} + \frac{1}{3} + \frac{1}{6}$$

Since every integer larger than 2 is either a multiple of 4 or a multiple of some odd prime, we just need to prove when $p > 3$ is a prime the above equation has a solution implies that it has a pairwise distinct solution. Assume we have positive integers x, y, z such that

$$4xyz = p(xy + yz + zx).$$

So $p|xyz$. Since p cannot divide all of x, y, z, without loss of generality, we assume either $(x, p) = 1$, $p|y$ and $p|z$ or $(x, p) = (y, p) = 1$ and $p|z$. In the former case, if

$y \neq z$ then we are done. Otherwise, we assume $y = z = tp$ and we have

$$\frac{4x - p}{x} = \frac{2}{t}.$$

Since $4x - p$ is odd, we have $2|t$. Let $t = 2t_1$, so $\frac{4}{p} = \frac{1}{x} + \frac{1}{t_1 p}$. If $t_1 = 1$, then $p = 3x$ which contradicts our assumption. Hence $t_1 > 1$ and we have

$$\frac{4}{p} = \frac{1}{x} + \frac{1}{(t_1+1)p} + \frac{1}{t_1(t_1+1)p}.$$

Notice that the denominators on the right are indeed pairwise distinct.

In the latter case, if $x \neq y$ then we are done. So suppose $x = y$ and $z = tp$, and we have $\frac{4}{p} = \frac{2}{x} + \frac{1}{tp}$. If $2|x$, then $x = 2s$, $s > 1$. Therefore

$$\frac{2}{x} = \frac{1}{s+1} + \frac{1}{s(s+1)}, s+1 \neq s(s+1).$$

If $x = 2s + 1$, then

$$\frac{2}{x} = \frac{1}{s+1} + \frac{1}{(s+1)(2s+1)}, s+1 \neq (s+1)(2s+1).$$

In either equation, if one of the denominator on the right is equal to tp, then we are in the situation as in the former case, and we are done. Otherwise, if neither of the two denominators is equal to tp, we are still done. Hence this finishes the proof.

Remark A.1. This proposition was first proved by Z. Ke, Q. Sun and X. Zhang [22] in 1964. But the proof here was worked out independently by the author when he was doing his undergraduate thesis.

□

Appendix B

Some elementary results on soluble residue classes

We begin with Lemma 4.3, which, for completeness, we reproduce here.

Proposition B.1. *For positive integers a, n with $(a, n) = 1$, the equation $\frac{a}{n} = \frac{1}{x} + \frac{1}{y}$ has positive integer solutions if and only if there exist two factors s and t of n such that $s + t$ is divisible by a.*

The following corollaries follow immediately.

Corollary B.2. *If p is a prime and a is a positive integer with $(a, p) = 1$, then $\frac{a}{p}$ is the sum of two unit fractions if and only if $p \equiv -1 \pmod{a}$.*

Corollary B.3. *$\frac{4}{n}$ is the sum of two unit fractions if and only if n is even or n has at least one prime factor congruent to 3 modulo 4.*

Observing that in order to establish the Erdős-Straus-Schinzel conjecture, it suffices to show the conjecture holds for prime numbers n. Hence without loss of generality, we just need to deal with the case that n is a prime.

Now, we can apply Proposition B.1 to get a sufficient condition for the conjecture.

Proposition B.4. *For any positive integer n, if there exist a positive integer k and its two factors s, t such that*

$$sn + t \equiv 0 \pmod{ak - 1}$$

then the equation $\frac{a}{n} = \frac{1}{x} + \frac{1}{y} + \frac{1}{z}$ *has positive integer solutions.*

Proof. Since sn and t are divisors of kn and $ak-1|sn+t$, then by Proposition B.1 we know $\frac{ak-1}{kn} = \frac{1}{x} + \frac{1}{y}$ has positive integer solutions in x, y. Hence $\frac{a}{n} = \frac{1}{x} + \frac{1}{y} + \frac{1}{kn}$. □

This criterion has a convenient equivalent form.

Proposition B.5. *Given positive integers* a *and* k*, and any factor* b *of* k^2*, if a positive integer* n *satisfies* $n \equiv -ab \pmod{ak-1}$ *or* $n \equiv 0 \pmod{ak-1}$*, then the equation* $\frac{a}{n} = \frac{1}{x} + \frac{1}{y} + \frac{1}{z}$ *has positive integer solutions.*

Proof. In the case that $n \equiv -ab \pmod{ak-1}$, we have $b = s't$ with $s'|k$ and $t|k$. Let $k = ss'$, so we have

$$sn \equiv -asb \equiv -ass't \equiv -akt \equiv -t \pmod{ak-1}$$

and we are done by Proposition B.4.

If $n \equiv 0 \pmod{ak-1}$, then by Proposition B.1 we know $\frac{a}{n} = \frac{1}{x} + \frac{1}{y}$ is solvable, and hence $\frac{a}{n} = \frac{1}{x} + \frac{1}{2y} + \frac{1}{2y}$. □

From now on, for the sake of convenience, instead of saying that for a given n, the equation $\frac{a}{n} = \frac{1}{x} + \frac{1}{y} + \frac{1}{z}$ has positive integer solutions, we will simply say n is soluble or solvable (if in the context the value of a is clear).

We explore a classical result by Mordell which can be obtained easily from the above propositions.

Corollary B.6 (Mordell)**.** *When* $a = 4$*, except for* $n \equiv 1, 121, 169, 289, 361, 529$ mod 840*,* n *is always soluble.*

Proof. Since $\frac{4}{2} = 1 + \frac{1}{2} + \frac{1}{2}$, so $n \equiv 0, 2, 4, 6 \pmod 8$ are soluble. By Proposition B.5, we know $n \equiv -1 \pmod 4$ and $n \equiv -16k \equiv -2 \pmod{8k-1}$ are soluble, so $n \equiv 3, 5, 7 \pmod 8$ are soluble.

By setting $k = 1, 2, 4$ in Proposition B.5, we know that $0, 2 \pmod 3$, $0, 3, 5, 6 \pmod 7$ and $7, 11, 13, 14 \pmod{15}$ are soluble residue classes. Besides $\frac{4}{5} = \frac{1}{2} + \frac{1}{5} + \frac{1}{10}$, so we have $0, 2, 3 \pmod 5$ are soluble residue classes. Now the conclusion follows by applying Chinese remainder theorem to the systems of congruence equations

$$\begin{cases} n \equiv 1 \pmod 8 \\ n \equiv 1 \pmod 3 \\ n \equiv 1, 4 \pmod 5 \\ n \equiv 1, 2, 4 \pmod 7 \end{cases}$$

□

We see from this elementary result, that it is quite easy to sieve out many integers through the soluble residue classes in Proposition B.5. Nevertheless, it is not clear at all whether these congruent classes can cover all the natural numbers, in which case the Erdős-Straus-Schinzel conjecture would have been solved. But we should be careful, for that all the soluble residue classes obtained in Proposition B.5 are quadratic non-residue classes, which means that squares can never be sieved out by this argument. Fortunately, as remarked before, it suffices to deal with primes. So eventually the conjecture can be solved by showing that all the primes are covered by soluble residue classes. However, very little is known about infinite covering congruent systems.

Proposition B.7. *Let b be a divisor of k^2, then $-4b$ is a quadratic non-residue modulo $4k-1$.*

Proof. We invoke the Jacobi symbol $\left(\frac{\cdot}{\cdot}\right)$ which is a natural generalization of the Legrend symbol to composite modulus. Similar to the Gauss quadratic reciprocity, the Jacobi symbol also admits the following reciprocity law:

For odd numbers a and b, we have

(i) $\left(\frac{-1}{a}\right) = (-1)^{\frac{a-1}{2}}$

(ii) $\left(\frac{2}{a}\right) = (-1)^{\frac{a^2-1}{8}}$

(iii) $\left(\frac{a}{b}\right) = (-1)^{\frac{a-1}{2}\frac{b-1}{2}} \left(\frac{b}{a}\right)$

If the Jacobi symbol is equal to 1 it does not necessarily catch quadratic residue but if instead it is equal to -1 it certainly picks up quadratic non-residue.

Now let $b = 2^\alpha c$, with c odd. So

$$\left(\frac{-4b}{4k-1}\right) = \left(\frac{-b}{4k-1}\right) = \left(\frac{-1}{4k-1}\right)\left(\frac{2}{4k-1}\right)^\alpha \left(\frac{c}{4k-1}\right).$$

Hitherto we have

$$\left(\frac{-1}{4k-1}\right) = (-1)^{\frac{4k-2}{2}} = (-1)^{2k-1} = -1$$

and

$$\left(\frac{2}{4k-1}\right)^{\alpha} = (-1)^{\frac{16k^2-8k}{8}\alpha} = (-1)^{k(2k-1)\alpha}$$

and

$$\left(\frac{c}{4k-1}\right) = (-1)^{\frac{c-1}{2}\frac{4k-2}{2}}\left(\frac{4k-1}{c}\right) = (-1)^{\frac{c-1}{2}}\left(\frac{-1}{c}\right) = (-1)^{\frac{c-1}{2}+\frac{c+1}{2}} = 1.$$

Also notice that if k is odd then $\alpha = 0$, hence $(-1)^{k(2k-1)\alpha} = 1$ regardless of the parity of k. Therefore, we conclude that

$$\left(\frac{-4b}{4k-1}\right) = -1$$

i.e. $-4b$ is a quadratic non-residue modulo $4k-1$. □

Remark B.1. Result of similar flavor has been established by Schinzel [31]. The same conclusion was proved in Elsholtz and Tao [7], but this proof was known to the author when he was an undergraduate.

Remark B.2. The original Erdős-Straus-Schinzel conjecture requires that solutions be positive integers. What if we relax them to be just integers? In 1978 Straus and Subbarao [33] proved that when $a < 40$, the equation $\frac{a}{n} = \frac{1}{x} + \frac{1}{y} + \frac{1}{z}$ has integer solutions for sufficiently large n. Their proof is somewhat tricky and it is not clear how to get it to work for all a.

We have seen that there are many soluble residue classes for the conjecture. Here is a result that is quantitative in nature.

Proposition B.8 (Vaughan, 1970)**.** *For each modulus $ak-1$, there are at least $\frac{1}{2}d(k^2)$ distinct soluble residue classes.*

Proof. Notice that k^2 has $\left[\frac{1}{2}d(k^2)\right]$ factors that are less than k, no pair among which are congruent to each other modulo $ak-1$. Now the desired result follows from Proposition B.5. □

In fact, instead of writing $\frac{1}{2}d(k^2)$, Vaughan wrote $\frac{1}{2}\sum_{t|k}|\mu(t)|d(\frac{k}{t})$, where $\mu(\cdot)$ is the Mobius function. But it turns out that Vaughan's version coincides with the version stated here, by virtue of $d(k^2) = \sum_{t|k}|\mu(t)|d(\frac{k}{t})$, which is a good exercise in elementary number theory. This proposition is the starting point of Vaughan's 1970 paper [38], in which he obtained by far the best upper bound estimate of $E_{a,3}(N)$. One could possibly improve the above result a little bit. Actually we have the following conjecture:

Conjecture B.9. *When k is not divisible by a, no pair of distinct factors of k^2 are congruent to each other modulo $ak-1$.*

Combining this conjecture with Proposition B.5, we immediately see that when k is not divisible by a there are $d(k^2)$ distinct soluble residue classes module $ak-1$. Also we remember that this is the largest possible number we can get through this argument. It is plausible that we am just missing some simple argument to prove the above conjecture. Anyway, for $a = 4$ we have tested it for $k \leq 10^8$, which strongly suggests its validity. Even with the conjecture, it is not possible to go much further along this approach. After all, Vaughan's version is good enough to obtain the best upper bound so far for $E_{a,3}(N)$.

Bibliography

[1] Cheng C.-X.; Dietel B.; Herblot M.; Huang J.-J.; Krieger H.; Marques D.; Mason J.; Mereb M.; Wilson S., *Some consequences of Schanuel's conjecture*, J. Number Theory 129 (2009), no. 6, 1464-1467.

[2] Croot E.; Dobbs D.; Friedlander J.; Hetzel A.; Pappalardi F., *Binary Egyptian fractions*, J. Number Theory 84(2000), no. 1, 63-79.

[3] Delange H., *Généralisation du Théorème de Ikehara*, Ann. Sci. Ecole Norm. Sup. (3) 71, 1954, 213-242.

[4] Elsholtz C., *The Erdős-Straus conjecture on* $\frac{4}{n} = \frac{1}{x} + \frac{1}{y} + \frac{1}{z}$, Diploma thesis, Technische Universität Darmstadt, 1996.

[5] Elsholtz C., *Sums of k unit fractions*, Ph.D. dissertation, Technische Universität Darmstadt, 1998.

[6] Elsholtz C., *Sums of k unit fractions*, Transactions of the American Mathematical Society, 2001, 353 (8): 3209-3227.

[7] Elsholtz C. and Tao T., *Counting the number of solutions to the Erdős-Straus equation on unit fractions*, arXiv:1107.1010.

[8] Guy R.K., *Unsolved problems in Number Theory*, second edition. Springer-Verlag, 1994.

[9] Halberstam H. and Roth K.F., *Sequences*, second edition, Springer-Verlag, New York-Berlin, 1983.

[10] Hardy G.H., *Ramanujian*, third edition, Chelsea Pub Co, 1978.

[11] Heath-Brown D.R., *Mean values of the zeta-function and divisor problems*, Recent Progress in Analytic Number Theory, Vol. 1 (Durham, 1979), Acad. Press, London-New York, 1981, pp. 115-119.

[12] Heath-Brown D.R., *The density of rational points on Cayleys cubic surface*, Proceedings of the session in analytic number theory and Diophantine equations, Bonner Math. Schriften 360 (2003).

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

[14] Huang J.-J., *A mean value theorem for the Diophantine equation* $axy-x-y=n$, Acta Mathematica Hungarica 134 (2012), no. 1, 68-78.

[15] Huang J.-J., *A mean value theorem for ternary Egyptian fractions*, preprint.

[16] Huang J.-J.; Marques D.; Mereb M., *Algebraic values of transcendental functions at algebraic points*, Bull. Aust. Math. Soc. 82 (2010), no. 2, 322-327.

[17] Huang J.-J. and Vaughan R.C., *Mean value theorems for binary Egyptian fractions*, J. Number Theory 131 (2011), no. 11, 1641-1656.

[18] Huang J.-J. and Vaughan R.C., *Mean value theorems for binary Egyptian fractions II*, to appear in Acta Arith., arXiv:1109.2274.

[19] Huang J.-J. and Vaughan R.C., *On the exceptional set for binary Egyptian fractions*, submitted.

[20] Ingham A.E., *Mean-value theorems in the theory of the Riemann zeta function*, Proc. London Math. Soc. (2) 27 (1926) 273-300.

[21] Ivić A., *The Theory of the Riemann Zeta-Function with Applications*. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1985. xvi+517 pp.

[22] Ke Z.; Sun Q.; Zhang X., *On the equation* $\frac{4}{n}=\frac{1}{x}+\frac{1}{y}+\frac{1}{z}$, Sichuan Daxue Xuebao, 1964, 3: 23-37.

[23] Landau E., *Handbuch von der Lehre der Verteilung der Primzahlen*, Teubner, Leipzig, 1909.

[24] Meurman T., *The mean twelfth power of Dirichlet L-functions on the critical line*, Ann. Acad. Sci. Fenn. Ser. A I Math. Dissertationes No. 52 (1984), 44 pp.

[25] Meurman T., *A generalization of Atkinson's formula to L-functions*, Acta Arith. 47 (1986), no. 4, 351–370.

[26] Misiurewicz M., *Ungelöste Probleme*, Elem. Math., 21(1966) 90.

[27] Montgomery H.L., *Topics in Multiplicative Number Theory*, Springer-Verlag, 1971.

[28] Montgomery H.L. and Vaughan R.C., *Multiplicative Number Theory I. Classical Theory*, Cambridge University Press, 2007.

[29] Nakayama M., *On the decomposition of a rational number into "stammbrüche"*, the Tôhuku Mathematical Journal, 1940, 46: 1-21.

[30] Pan C.D. and Pan C.B., *Foundation to Analytic Number Theory (Chinese)*, Science Press, Beijing, 1991.

[31] Schinzel A., *On sums of three unit fractions with polynomial denominators*, Funct. Approx. Comment. Math., 2000, 28: 187-194.

[32] Shan Z., *On the diophantine equation* $\sum_{i=0}^{k} \frac{1}{x_i} = \frac{a}{n}$, Chinese Annals of Mathematics, 1986, 7B: 213-220.

[33] Straus E.G. and Subbarao M.V., *On the representation of fractions as sum and difference of three simple fractions*, Proceedings of the Seventh Manitoba Conference on Numerical Mathematics and Computing, 561-579, Utilitas Mathematica Publishing Inc., 1978.

[34] Swett A., *The Erdos-Strauss conjecture*, http://math.uindy.edu/swett/esc.htm, 1999.

[35] Tenenbaum G., *Introduction to analytic and probabilistic number theory*, Cambridge University Press, 1995.

[36] Titchmarsh E.C., *The Riemann Zeta-Function, 2nd edition, revised by D.R. Heath-Brown*, Oxford, 1986.

[37] Vaughan R.C., *An application of the large sieve to a diophantine equation*, Berichte aus dem Mathematischen Forschungsinstitut Oberwolfach, Heft 5, Bibliographisches Institut, March 1970: 203-207.

[38] Vaughan R.C., *On a problem of Erdős, Straus and Schinzel*, Mathematika, 1970, 17: 193-198.

[39] Viola C., *On the diophantine equation* $\prod_{i=0}^{k} x_i - \sum_{i=0}^{k} x_i = n$ *and* $\sum_{i=0}^{k} \frac{1}{x_i} = \frac{a}{n}$, Acta Arithmetica, 1973, 22: 339-352.

[40] Webb W.A., *On* $\frac{4}{n} = \frac{1}{x} + \frac{1}{y} + \frac{1}{z}$, Proc. Amer. Math. Soc., 1970, 25: 578-584.

[41] Webb W.A., *Rationals not expressible as a sum of three unit fractions*, Elem. Math., 1974, 29: 1-6.

Vita

Jing-Jing Huang

Jing-Jing Huang was born in Jiujiang, Jiangxi province in China in 1985, and grew up there. He was attracted by Mathematics when he came across an expository number theory book in the junior middle school. He won the First Prize in the National Mathematical Competition in 2002 and was hence recommended for admission to the Nankai University where he was introduced to higher mathematics and enjoyed wonderful four years. He particularly benefits from those analysis and algebra courses taught by Li Jun, Meng Daoji, Zhou Xingwei and Zhang Lun. Though without number theory courses taught at Nankai, he studied number theory by himself and became deeply fond of the subject. He received his B.Sc. Math. degree in 2007 after which he came to United States to pursue his Ph.D. at the Pennsylvania State University.

Shortly after his arrival at Penn State in 2007, Huang joined the Number Theory group and became a graduate student of Robert Vaughan and Winnie Li. His first research paper was published in 2009 and subsequently he accomplished six more during his thesis study.

While completing his thesis research, Huang was awarded the Pritchard Dissertation Fellowship by the Department of Mathematics.

For the time being, Huang's research interests are in number theory, Diophantine problems, Hardy-Littlewood methods, automorphic forms and L-functions.

The majority of the results in this thesis are based on Huang's papers: [14], [15], [17], [18] and [19].