ARITHMETIC BIASES FOR BINARY QUADRATIC FORMS

JEREMY SCHLITT

A Thesis
in
The Department
of
Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Science (Mathematics) at
Concordia University
Montreal, Quebec, Canada

August 2023

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Jeremy Schlitt

Entitled: Arithmetic Biases for Binary Quadratic Forms
and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Mathematics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality. Signed by the final Examining Committee:

______________________________ Thesis Supervisor
Dr. C. David

______________________________ Examiner
Dr. G. Rosso

______________________________ Examiner
Dr. C. Pagano

Approved by ______________________________
Dr. Y. Chaubey (Graduate Program Director)

____
Dr. P. Sicotte (Dean of Faculty)

____
Date

Abstract for MSc

Arithmetic Biases for Binary Quadratic Forms

Jeremy Schlitt

Concordia University, 2023

The prime number theorem for arithmetic progressions tells us that there are asymptotically as many primes congruent to 1 mod 4 as there are congruent to 3 mod 4. That being said, Chebyshev noticed that (numerically) there almost always seems to be slightly more primes congruent to 3. This simple fact has a highly non-trivial explanation. Rubinstein and Sarnak [RS94] proved that the assumption of some natural (yet still unproven) conjectures, there is a way to prove that there are more primes congruent to 3 than congruent to 1 more than half of the time (in an appropriate sense).

Many other sets of integers demonstrate a bias towards a certain residue class modulo some number q. Recently, Gorodetsky [Gor22] showed that the sums of two squares exhibit a Chebyshev-type bias, and that in this case the conjectures one must assume to prove the existence of the bias are weaker. In this thesis, we present two papers (chapters 2 and 3) which demonstrate some bias in arithmetic progressions for sets of integers that are represented by a given binary quadratic form.

In chapter 2, we examine a bias towards the zero residue class for the integers represented by binary quadratic forms. In many cases, we are able to prove that the bias comes from a secondary term in the associated asymptotic expansion (unlike Chebyshev's bias, which lives somewhere at the level of $O(x^{1/2+\epsilon})$.) In some other cases, we are unable to prove that a bias exists, even though it is present numerically. We then make a conjecture on the general situation which includes the cases we could not prove. Many interesting results on the distribution of the integers represented by a quadratic form are proven, and the paper finishes with some numerical data that is illustrative of the generic data for any quadratic form.

In chapter 3, we examine a different kind of bias. We ask for the distribution of pairs of sums of two squares in arithmetic progressions, i.e. how many numbers are the sum of two squares, congruent to a mod q, and are such that the next largest sum of two squares is congruent to b mod q. We prove that when $q \equiv 1$ mod 4, we have equidistribution among the q^2 possible pairs of residue classes. That being said, there exist bizarre numerical biases, most notably a negative bias towards repetition. The main purpose of the second paper is to provide a conjecture which explains the bias, via a secondary and tertiary term in the associated asymptotic expansion. We then support this conjecture with both numerical and theoretical evidence. The paper contains many partial results in the direction of the conjecture, as well as some theorems on the sums of two squares that are of independent interest. For example, we provide an integral representation for the number of integers not exceeding x which are the sum of two squares. This integral representation is akin to $li(x)$ for primes, in that it has a $O(x^{1/2+\epsilon})$ error term under the Generalized Riemann Hypothesis.

Acknowledgements

I am eternally grateful to my mentor and thesis supervisor, Professor Chantal David. In addition to having a brilliant mathematical mind, Chantal is a kind and caring supervisor whom I am privileged to have met. I am especially appreciative of the autonomy she gave me to choose my own topic, and her constant enthusiasm about said topic.

I would like to thank Professors Giovanni Rosso and Carlo Pagano for serving on my defence committee, and for the lovely and insightful courses they taught during my time as a student at Concordia.

I also would like to thank Lucile Devin and Jungbae Nam for helping to make my first research project enjoyable and profitable.

I wish to thank my family for their constant support. Mom, Justin, EO, I hope I make you all proud. Thanks also to Kazi, my brother from another mother, for his wisdom, and all the laughs.

Last, but certainly not least, I wish to thank my partner Stephanie Tanasia for reminding me to enjoy life. Stephanie's warm and nurturing nature helped me get through the most stressful moments of my time as a master's student. It is no exaggeration to say that this thesis would not have been possible without her, and I will never stop finding ways to express my gratitude towards her. Aku sayang kamu banget!

Contribution of Authors

Chapter 1 is an original contribution of the author.
Chapter 2 is an original contribution of the author, which will be submitted to an academic journal in 2023.
Chapter 3 is an original contribution of Chantal David, Lucile Devin, Jungbae Nam and the author. It was published in the Mathematische Annalen journal on November 25^{th} 2021.

All authors reviewed the final manuscript and approved of the contents.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 An Overview of this Thesis

This thesis is comprised of two research papers in analytic number theory. The first paper is a complete manuscript which will be submitted to a journal in 2023. The second paper is collaborative work with Chantal David, Lucile Devin, and Jungbae Nam which was completed in 2020-2021 and published in Mathematische Annalen in November 2021. Both papers are, in some respect, related to the distribution in arithmetic progressions of integers represented by a given quadratic form. A positive definite binary quadratic form is one that can be expressed in as $ax^2 + bxy + cy^2$, where a, b, and c are integers, $a > 0$, and the discriminant $D = b^2 - 4ac$ is negative. The values of the quadratic form are strictly positive except when $\mathrm{x} = \mathrm{y} = 0$. Positive definite binary quadratic forms are important objects in number theory and have numerous applications, including the study of quadratic fields, the distribution of primes, and the representation of integers by such forms.

Binary quadratic forms have a long and rich history in mathematics that spans over two millennia. The first known reference to these forms dates back to Diophantus of Alexandria in the 3rd century AD, who studied equations of the form $ax^2 + bx = cy^2$. However, it was not until the work of Indian mathematicians, including Aryabhata and Brahmagupta, in the 5th and 6th centuries AD that a systematic theory of binary quadratic forms began to emerge. In the 18th and 19th centuries, European mathematicians, such as Lagrange, Gauss, and Dirichlet, further developed the theory of binary quadratic forms and made significant contributions to their applications in number theory. Today, binary quadratic forms continue to be an active area of research in mathematics, with ongoing investigations into their properties, applications, and connections to other fields.

A seemingly innocuous question one can ask about the integers represented by a given binary quadratic form is: How are these numbers distributed modulo some fixed integer? Prachar [Pra53] proved a general theorem about the asymptotic proportion of integers represented by the quadratic form $x^2 + y^2$ that lie in each residue class modulo q. One consequence of Prachar's theorem is that there is equidistribution modulo 5 for sums of two squares; one fifth of the numbers not exceeding X are of the form $x^2 + y^2$ are congruent to a mod 5 for each residue class $a \pmod 5$, as X tends to infinity.

By examining numerical data (section 2.5), one quickly sees that there is something intriguing happening with the 0 residue class. It seems that there are more sums of two squares congruent to 0 mod 5 than there are congruent to anything else. In the first paper (chapter 2), this phenomenon is explored in the more general setting where we replace 5 with some prime q, and we replace $x^2 + y^2$ with some positive definite integral binary quadratic form $ax^2 + bxy + cy^2$, subject to some conditions which will be explained later.

In this first paper, we observe that a bias similar to that we observe for the sums of two squares modulo 5 occurs for pretty much any form modulo any prime. In particular, the zero residue class modulo q always contains more integers than any other class. We prove that for a large family of quadratic forms, this bias arises from a secondary term in an associated asymptotic expansion, and we give quantitative characterization of the bias. We also extend this claim to an even more general family of quadratic forms, but our results there are more qualitative in nature. Finally, in the first paper, we make a conjecture on the general situation for all (suitable) integral binary quadratic forms, and provide numerical evidence to support our conjecture.

The second paper (chapter 3), inspired by the work of Lemke Oliver and Soundararajan [LOS16], studies a different kind of arithmetic progression bias. In particular, we know that asymptotically 1/5 of the sums of two squares are a mod 5 for any a. We wish to know, asymptotically, the number of sums of two squares that are a mod 5 *and* are such that the next largest sum of two squares is b mod 5 for each pair a,b. It is shown that asymptotically each pair represents 1/25 of the total count, but there are perplexing irregularities in numerical data. In particular, there is a bias against repetitions, meaning that if a sum of two squares is, say, 0 mod 5, it is (at least according to numerical data) less likely that the next largest sum of two squares will also be 0 mod 5.

The second paper gives a conjectural explanation for this negative bias towards repetitions, and also explains some other phenomena one can observe in the data. These conjectures are themselves based on widely believed conjectures, namely the Generalized Riemann Hypothesis (GRH), as well as a version of Hardy and Littlewood's pair correlation conjecture for sums of two squares. In addition to using these conjectures to formulate our own, we also prove partial results in the direction of the pair correlation conjecture. Overall, the paper contains many original theorems on sums of two squares that are of interest in their own right, such as an integral representation for the number of sums of two squares up to x with a $O(x^{1/2+\epsilon})$ error term under GRH.

The commonality between both the first and second papers is that there exists some kind of numerical bias in arithmetic progressions for a set of integers (or tuples of integers) being represented by some quadratic form, and that this bias is explained by the computation of secondary order terms in the associated asymptotic expansions. This being said, the results and the methods used in both papers differ vastly. The first paper relies almost entirely on classical techniques in analytic number theory, whereas the second paper also combines some basic ideas from combinatorics, such as the inclusion-exclusion principle.

Chapter 2

Biases Towards the Zero Residue Class for Quadratic Forms in Arithmetic Progressions

2.1 Introduction

A well known theorem of Landau [Lan09] gives an asymptotic estimate on the number of positive integers not exceeding x that are the sum of two squares. Some years after Landau's original proof, Prachar ([Pra56]) proved a generalization to arithmetic progressions. Precisely, he showed that for fixed integers a,q, with $(a,q) = 1$, and $a \equiv 1 \bmod (4,q)$, as $x \to \infty$,

$$\sum_{\substack{n \leq x \\ n = \square + \square \\ n \equiv a \bmod q}} 1 \sim K_q \frac{x}{(\log x)^{1/2}}, \tag{2.1}$$

where

$$K_q = K \frac{(4,q)}{(2,q)q} \prod_{\substack{p \mid q \\ p \equiv -1 \bmod 4}} \left(1 + \frac{1}{p}\right),$$

and K is the Landau-Ramanujan constant, defined by

$$K = \frac{1}{\sqrt{2}} \prod_{p \equiv -1 \bmod 4} \left(1 - \frac{1}{p^2}\right)^{-1/2}.$$

The case $q = 1$ of course corresponds with Landau's theorem. See [Iwa76] for a more modern account of Prachar's theorem and some of its improvements.

In the present article, we are interested in the distribution in arithmetic progressions of integers represented by a given binary quadratic form. We will denote by $B_f(x;q,a)$ the number of integers not exceeding x that are congruent to a modulo q, and represented by the form f. In what follows, D will always denote a negative fundamental discriminant. As a first result, we have an extension of Prachar's theorem[1] to forms other than x^2+y^2.

Theorem 2.1.1. *Let f be a binary quadratic form of discriminant D, and let q be a positive integer such that $(q,2D)=1$. One has, for all a satisfying $(a,q)=1$:*

$$B_f(x;q,a) = \frac{1}{q} \prod_{\substack{p|q \\ \left(\frac{D}{p}\right)=-1}} \left(1+\frac{1}{p}\right) B_f(x) + \mathcal{O}\left(\frac{x}{(\log x)^{2/3}}\right),$$

where $B_f(x) = B_f(x,1,1)$ denotes the number of integers not exceeding x which are represented by the form f. In particular, when q is a prime number, one has

$$B_f(x;q,a) = c(q,a)B_f(x) + \mathcal{O}\left(\frac{x}{(\log x)^{2/3}}\right),$$

where

$$c(q,a) := \begin{cases} \frac{1}{q} & \text{if } \left(\frac{D}{q}\right) = 1 \\ \frac{q+1}{q^2} & \text{if } \left(\frac{D}{q}\right) = -1, a \not\equiv 0 \bmod q \\ \frac{1}{q^2} & \text{if } \left(\frac{D}{q}\right) = -1, a \equiv 0 \bmod q \end{cases}$$

The computation of the main term for $B_f(x)$ is due to Bernays [Ber12]. See Section 2.1 of [MO06] for an excellent exposition on Bernay's theorem, due to Moree and Osburn.

As we will see in the proof of Theorem 2.1.1, it is non-trivial to make the leap from x^2+y^2 to any binary quadratic form of negative fundamental discriminant, as one must make considerations about the ideal class group of $\mathbb{Q}(\sqrt{D})$, which in general may not be trivial as it is when $D=-4$.

Theorem 2.1.1 tells us what we should expect as the proportion of integers that lie in each nonzero residue class, and when q is prime, this implies the proportion for the zero residue class. However, a look at some of the tables of Section 2.5 giving values of $B_f(x;q,a)$ for varying f, q, a suggests an interesting discrepancy between Theorem 2.1.1 and actual data. It appears that there is a numerical bias towards the 0 residue class in all examples. In cases when $\left(\frac{D}{q}\right)=1$, we expect a proportion of $1/q$ in each class, and yet we observe that the 0 class contains more integers than the non-zero classes in an apparent way. Similarly, in cases when $\left(\frac{D}{q}\right)=-1$, we expect there to be roughly $1/q^2$ of the integers falling into the zero residue class, but in actuality it seems like this estimate is always an under-count. We give in this paper a theoretical explanation for this bias in some cases. Hereafter, $C(D)$ denotes the group of reduced forms of discriminant D with $h := |C(D)|$, and $G(D)$ denotes the genus group; $G(D) \cong C(D)/C(D)^2$. In the case

[1] Under slightly stricter assumptions about the modulus.

when there is a single quadratic form per genus in the genus group, i.e. $C(D) \cong G(D)$, we explicitly compute the second term in the asymptotic expansion of $B_f(x;q,a)$, which accounts for the bias:

Theorem 2.1.2. *Let D be a fundamental discriminant such that $C(D) \cong G(D)$, and let $f \in C(D)$ be a given form. Let q be a prime modulus for which $(q,2D) = 1$. One has*

$$B_f(x;q,a) = c(q,a)\left(a_0\frac{x}{(\log x)^{1/2}} + a_1\Big(1 - \frac{a_0}{a_1}\delta(a,q)\Big)\frac{x}{(\log x)^{3/2}}\right) + O\left(\frac{x}{(\log x)^{5/2}}\right),$$

where $c(q,a)$ is as in Theorem 2.1.1, a_0 and a_1 are defined by (2.2), *and*

$$\delta(q,a) = \begin{cases} \frac{\log q}{2(q-1)}, & \text{if } \left(\frac{D}{q}\right) = 1, \quad a \not\equiv 0 \bmod q \\ -\frac{\log q}{2}, & \text{if } \left(\frac{D}{q}\right) = 1, \quad a \equiv 0 \bmod q \\ \frac{\log q}{q-1}, & \text{if } \left(\frac{D}{q}\right) = -1, \quad a \not\equiv 0 \bmod q \\ -\log q, & \text{if } \left(\frac{D}{q}\right) = -1, \quad a \equiv 0 \bmod q. \end{cases}$$

Theorem 2.1.2 tells us that the secondary term of $B_f(x;q,a)$ is generally larger when $a \equiv 0 \bmod q$, as opposed to when $a \not\equiv 0 \bmod q$, explaining the numerical bias.

Let us write $1_R(n)$ to be the function that equals 1 if n is represented by a form in the genus $R \in G(D)$, and $1_R(n) = 0$ otherwise. Let us also define

$$B_R(x;q,a) := \sum_{\substack{n \le x \\ n \equiv a \bmod q}} 1_R(n).$$

The proofs of Theorems 2.1.1 and 2.1.2 both rely on the fact that one can compute $B_R(x;q,a)$, by writing a generating series with genus group characters. The generating series we obtain will have an essential singularity at $s = 1$, which allows us to apply the Landau-Selberg-Delange (LSD) method. This process will immediately yield Theorem 2.1.2, as we have by assumption in this case that $B_f(x;q,a) = B_R(x;q,a)$. To conclude the proof of Theorem 2.1.1 once we have the main term of $B_R(x;q,a)$ in hand, we apply the following strong result of Fomenko ([Fom98]):

Theorem 2.1.3. *Let f be a form in the genus R. We have (as $x \to \infty$)*

$$B_f(x;q,a) = B_R(x;q,a) + ‰\left(\frac{x}{(\log x)^{2/3}}\right).$$

In an appropriate sense, Theorem 2.1.3 tells us that almost all integers represented by *some* form in the genus R are actually represented by *all* forms in that genus.

Turning now to the case when there is a single genus of forms ($G(D) \cong \{1\}$), but with several forms in that genus, we prove some partial results which give good evidence to there being a numerical bias towards the zero residue class modulo q. Rather astonishingly, though the secondary term in this case is of a different nature than the secondary term of Theorem 2.1.2, it still espouses a bias towards the zero residue class.

Theorem 2.1.4. *Let D be a discriminant for which $C(D)$ is cyclic of odd order h. Let $[f^\star]$ be a generator of $C(D)$, and $H = \langle [f^\star]^{p_0} \rangle$, where p_0 is the smallest prime divisor of h. Let q be an integer such that $(q,2D) = 1$, and let $B'_f(x;q,a)$ denote the number of squarefree integers not exceeding x that are coprime to $2D$, are represented by the form f, and are congruent to a mod q. Let $A_1(f,q,a)$ be the constant such that (as $x \to \infty$)*

$$B'_f(x;q,a) \sim A_1(f,q,a)\frac{x}{(\log x)^{1/2}}.$$

Then, one has, for every a such that $(a,q) = 1$,

$$B'_f(x;q,a) = A_1(f,q,a)\frac{x}{(\log x)^{1/2}} - c'(q,a)A_2(f)\frac{x}{(\log x)^{1-1/(2p_0)}}(\log\log x)^r(1+o(1)),$$

where $A_2(f)$ is a positive constant depending only on f,

$$r = \begin{cases} p_0 - 2, & [f] \in H, \\ p_0 - 3, & \textit{otherwise}, \end{cases}$$

$$c'(q,a) = \frac{1}{\phi(q)} \prod_{p|q} \left(1 + \frac{\nu_H(p)}{p}\right)^{-1},$$

and $\nu_H(p)$ is defined to be 1 if p is represented by a class of forms in H, and $\nu_H(p) = 0$ otherwise. In particular, if q is taken to be prime, then

$$c'(q,a) = \begin{cases} \frac{1}{q+1}, & a \equiv 0 \bmod q \ \textit{and}\ \nu_H(q) = 1 \\ \frac{q}{q^2-1}, & \textit{if}\ a \not\equiv 0 \bmod q \ \textit{and}\ \nu_H(q) = 1 \\ \frac{1}{q-1}, & a \equiv 0 \bmod q \ \textit{and}\ \nu_H(q) = 0 \\ o(1), & \textit{if}\ a \not\equiv 0 \bmod q \ \textit{and}\ \nu_H(q) = 0. \end{cases}$$

Note that $q/(q^2-1) > 1/(q+1)$, so that (for q prime) $-c'(q,a)$ is larger when $a \equiv 0$ and smaller otherwise.

An examination of numerical data suggests that the bias still exists when one removes the restriction that the integers should be squarefree and coprime to $2D$. See Table 2.8. In Section 2.3.2, we will discuss why dropping the squarefreeness condition makes the implied constants in-explicit.

Summary of main theorems: When $G(D) \cong C(D)$, then the biased secondary term will be of size $x/(\log x)^{3/2}$, and will arise from a result on the equidistribution of arithmetic progressions for integers represented by a given genus (Theorem 2.2.1).

When $C(D)$ is cyclic and of odd order, then the biased secondary term will be of size $x(\log\log x)^{\beta}/(\log x)^{\alpha}$, for some $\beta \in \mathbb{Z}_{\geq 0}$, and $\alpha \in [5/6,1)$. The secondary term in this case arises from a characterization of the integers which are represented by some form in the genus of f *but not* by f itself, which will generalize the work of [Gol01].

Several recent works have been concerned with some aspects of the behaviour of the form $x^2 + y^2$ in arithmetic progressions. In [Gor22], Gorodetsky demonstrates a Chebyshev-type bias towards quadratic residues for numbers of the form $x^2 + y^2$. Most notable of Gorodetsky's result is that it occurs in a *natural density* sense, unlike the logarithmic density one obtains when examining Chebyshev's bias for primes. This distinction allows Gorodetsky to loosen the hypothesis on the linear independence of zeros of ζ and other associated L-functions.

Inspired by the earlier work of Lemke Oliver and Soundararajan [LOS16], David, Devin, Nam, and the author studied the distribution of consecutive sums of two squares in arithmetic progressions in [DDNS21]. Theorem 2.4 of that paper is a special case of Theorem 2.1.2 in the present paper.

The function field analogue of Landau's theorem has also been explored. In [GR21], Gorodetsky and Rogers explore the variance of sums of two squares in short intervals for $\mathbb{F}_q[T]$. Theorem B.1 of their article is of interest, as it provides an integral representation for the number of integers that are the sum of two squares not exceeding x, with a $O(x^{1/2+\epsilon})$ error term (under GRH). It would be interesting to consider the extension of this theorem to more general families of binary quadratic forms.

2.2 Proof of Theorems 1 and 2

2.2.1 Proof of Theorem 1

Theorem 2.1.3 tells us that the main term in the asymptotic expansion of $B_R(x;q,a)$ matches that of $B_f(x;q,a)$. In light of this fact, it can be seen that Theorem 2.1.1 is implied from the following:

Theorem 2.2.1. *Let q be a positive integer, and let $(q,2D) = 1$, where D is a negative fundamental discriminant. Let $R \in G(D)$ be a genus of forms. Then, for every $J \in \mathbb{N}$, we can write*

$$B_R(x) := B_R(x;1,1) = \sum_{j=0}^{J} \frac{a_j x}{(\log x)^{1/2+j}} + O\left(\frac{x}{(\log x)^{3/2+J}}\right) \tag{2.2}$$

and for every a satisfying $(a,q) = 1$, we can write

$$B_R(x;q,a) = \sum_{j=0}^{J} \frac{b_j x}{(\log x)^{1/2+j}} + O\left(\frac{x}{(\log x)^{3/2+J}}\right)$$

for positive constants a_j and b_j that do not depend on R or on a. a_j is defined by (2.9)*, and the first two b_j are given by*

$$b_0 = a_0 \frac{1}{q} \prod_{\substack{p|q \\ (D|p)=-1}} (1+p^{-1}), \tag{2.3}$$

$$b_1 = a_1 \frac{1}{q} \prod_{\substack{p|q \\ (D|p)=-1}} (1+p^{-1}) \left(1 - \frac{a_0}{2a_1} \left(\sum_{p|q} \frac{\log p}{p-1} - \sum_{\substack{p|q \\ (D|p)=-1}} \frac{\log p}{p+1} \right) \right). \tag{2.4}$$

Proof of (2.2.1). We will loosely follow the proof of a similar theorem which appears in [Lut67], where asymptotic for the number of rational integers that are the norm on an algebraic integer of a given quadratic number field is obtained. We begin with a Lemma which completely describes how the genus characters act on the primes.

Lemma 2.2.2. *Let D be the discriminant of the imaginary quadratic field K. There is a one-to-one correspondence between the characters of the genus group $G(D)$ and the factorizations $D = uv$ of D into two fundamental discriminants u and v (treating uv and vu as the same factorization). The relation between the genus characters $\psi \in \widehat{G(D)}$ and the factorizations is expressed by*

$$\psi(\mathfrak{p}) := \begin{cases} \left(\frac{u}{N(\mathfrak{p})}\right), & \text{if } (u,\mathfrak{p}) = 1, \\ \left(\frac{v}{N(\mathfrak{p})}\right), & \text{if } (v,\mathfrak{p}) = 1, \end{cases}$$

where $\mathfrak{p}$ is any prime ideal of $\mathcal{O}_K$

Proof. Note that any real character of the ideal class group $C(D)$ will correspond with a character of the genus group $G(D) = C(D)/C(D)^2$, as

$$\begin{aligned} \psi \text{ is a character of } G(D) &\iff \psi \text{ is a character of } C(D) \\ &\qquad \text{satisfying } \psi(g^2 n) = \psi(g'^2 n) \quad \forall g, g', n \in C(D) \\ &\iff \psi \text{ is a real character of } C(D). \end{aligned}$$

With this in mind, we know that the number of genera of discriminant D is $2^{\mu-1}$. The integer μ is defined as follows: if $D \equiv 1 \bmod 4$, $\mu = s$, the number of distinct odd prime factors of D. If $D = 4n$, then

$$\mu = \begin{cases} s, & n \equiv 1 \bmod 4 \\ s+1 & n \equiv 2,3 \bmod 4 \\ s+1 & n \equiv 4 \bmod 8 \\ s+2 & n \equiv 0 \bmod 8 \end{cases}$$

See Theorem 3.15 of [Cox13] for the proof of this fact. This proves that there are $2^{\mu-1}$ characters of $G(D)$. We note that we get one real character $\psi(\mathfrak{p})$ for each factorization $D = uv$ as defined in the statement of the lemma. One can calculate that there are in total $2^{\mu-1}$ such factorizations of D, so that the set of characters of $G(D)$ which are as in the statement of the lemma (coming from a factorization $D = uv$) are actually all the characters of $G(D)$. This proves the $1-1$ correspondence. □

From each genus character, we construct the generating series

$$F(s,\psi) := \sum_{n\geq 1} b(n,\psi)n^{-s},$$

for $\Re(s) > 1$, where

$$b(n,\psi) = \begin{cases} 0, & \textit{if } n \textit{ is not the norm of any ideal,} \\ \psi(\mathfrak{a}), & \textit{if } n = N(\mathfrak{a}). \end{cases}$$

If $p \in \mathbb{Z}$ is such that $(D|p) = 0$, then $pO_K = \mathfrak{p}^2$, and so $p = N(\mathfrak{p})$, if $(D|p) = 1$, then $pO_K = \mathfrak{p}_1\mathfrak{p}_2$, so that $p = N(\mathfrak{p}_1) = N(\mathfrak{p}_2)$. Finally, if $(D|p) = -1$, then $pO_K = \mathfrak{p}$, so that $p^2 = N(\mathfrak{p})$, and p is not the norm of any ideal. These remarks imply that $b(n,\psi)$ is well defined.

Our reason for introducing $b(n,\psi)$ is that it satisfies an orthogonality relation that allows us to detect when $1_R(n) = 1$. Indeed, when n is the norm of some ideal $\mathfrak{a}$, we have:

$$\begin{aligned} 1_R(n) &= \frac{1}{G(D)} \sum_{\psi\in\widehat{G(D)}} b(n,\psi)\psi(R)^{-1} \\ &= \frac{1}{G(D)} \sum_{\psi\in\widehat{G(D)}} \psi(\mathfrak{a})\psi(R)^{-1} \\ &= \begin{cases} 1, & [\mathfrak{a}] \in R \\ 0, & \textit{otherwise.} \end{cases} \end{aligned} \tag{2.5}$$

Above, $[\mathfrak{a}]$ refers to the ideal class of $\mathfrak{a}$ in the ideal class group $C(D)$. When n is not the norm of any ideal, then $1_R(n) = b(n,\psi) = 0$ for every $\psi \in \widehat{G(D)}$, so the first line of (2.5) still holds.

Using the multiplicativity of $b(n,\psi)$, we find that

$$F(s,\psi) = \prod_{p|D}(1-\psi(\mathfrak{p})p^{-s})^{-1} \prod_{(D|p)=1}(1-\psi(\mathfrak{p})p^{-s})^{-1} \prod_{(D|p)=-1}(1-\psi(\mathfrak{p})p^{-2s})^{-1}.$$

In each term of the above product, $\mathfrak{p}$ denotes any prime above p. In the cases when $pO_K = \mathfrak{p}_1\mathfrak{p}_2$, we note that $\psi(\mathfrak{p}_1) = \psi(\mathfrak{p}_2)$, so we can choose either one. We now explicitly write the Euler factors above a prime for the three possible cases (ramifies, splits, inert).

Ramifies: If $p|D$, then $p|uv$, and $p|u \iff p \nmid v$, and

$$(1-\psi(\mathfrak{p})p^{-s}) = \left(1-\left(\frac{u}{p}\right)p^{-s}\right)\left(1-\left(\frac{v}{p}\right)p^{-s}\right),$$

since one of the two factors on the right hand side is always 1.

Splits: If $(D|p)=1$, then $\left(\frac{u}{p}\right)\left(\frac{v}{p}\right)=1$, and

$$\begin{aligned}(1-\psi(\mathfrak{p})p^{-s}) &= \left(1-\left(\frac{u}{p}\right)p^{-s}\right)\\ &= \left(1-\left(\frac{v}{p}\right)p^{-s}\right)\\ &= \left(1-\left(\frac{u}{p}\right)p^{-s}\right)^{1/2}\left(1-\left(\frac{v}{p}\right)p^{-s}\right)^{1/2}.\end{aligned}$$

Inert: if $(D|p)=-1$, then $N(\mathfrak{p})=p^2$, hence $\left(\frac{u}{N(\mathfrak{p})}\right)=\left(\frac{v}{N(\mathfrak{p})}\right)=1$, and

$$(1-\psi(\mathfrak{p})p^{-2s}) = (1-p^{-2s}) = \left(1-\left(\frac{u}{p}\right)p^{-s}\right)\left(1-\left(\frac{v}{p}\right)p^{-s}\right).$$

Putting all this information together, we find that

$$\begin{aligned}F(s,\psi) = &\prod_{p|D}\left(1-\left(\frac{u}{p}\right)p^{-s}\right)^{-1}\left(1-\left(\frac{v}{p}\right)p^{-s}\right)^{-1}\\ &\times\prod_{(D|p)=1}\left(1-\left(\frac{u}{p}\right)p^{-s}\right)^{-1/2}\left(1-\left(\frac{v}{p}\right)p^{-s}\right)^{-1/2}\\ &\times\prod_{(D|p)=-1}\left(1-\left(\frac{u}{p}\right)p^{-s}\right)^{-1}\left(1-\left(\frac{v}{p}\right)p^{-s}\right)^{-1}.\end{aligned}$$

Let us define

$$\begin{aligned}L_u(s) &= \prod_p\left(1-\left(\frac{u}{p}\right)p^{-s}\right)^{-1}\\ L_v(s) &= \prod_p\left(1-\left(\frac{v}{p}\right)p^{-s}\right)^{-1}.\end{aligned}$$

We now have

$$F(s,\psi)^2 = L_u(s)L_v(s)A(s), \tag{2.6}$$

where

$$A(s)=\prod_{p|D}\left(1-\left(\frac{u}{p}\right)p^{-s}\right)^{-1}\left(1-\left(\frac{v}{p}\right)p^{-s}\right)^{-1}\prod_{(D|p)=-1}(1-p^{-2s})^{-1}.$$

We note that $A(s)$ is analytic for $\Re(s)>1/2$, and $L_u(s)L_v(s)$ is entire unless $u=1,v=D$, since when $u\neq 1, v\neq D$, both L_u and L_v can be viewed as Dirichlet L-functions of non-principal characters[2]. We note that the pair $u=1$,$v=D$ corresponds to the principal genus character ψ_0, and we have

$$L_1(s)L_D(s)=\zeta(s)L_D(s).$$

Taking (principal) square roots on either side of (2.6), we have

$$F(s,\psi_0)=\left(\zeta(s)A(s)L_D(s)\right)^{1/2}.$$

We wish to use the Dirichlet series $F(s,\psi_0)$ to make a conclusion about the partial sums $\sum_{n\leq x} b(n,\psi_0)$. We can make use of Perron's formula in the usual way to get the relation

$$\sum_{n\leq x} b(n,\psi_0)=\frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty}F(s,\psi_0)\frac{x^s}{s}ds,$$

but we are not able to apply Cauchy's theorem to evaluate the contribution from the singularity at $s=1$ of the integrand, since it is not a pole. Instead, we appeal to the LSD method:

Theorem 2.2.3. *[Kou19, Theorem 13.2] Let $f(n)$ be a multiplicative function with generating function $F(s)=\sum_{n\geq 1} f(n)n^{-s}$. Suppose there exists $\kappa\in\mathbb{C}$ be such that for x large enough*

$$\sum_{p\leq x} f(p)\log p=\kappa x+O_A\left(x/(\log x)^A\right),$$

for each fixed $A>0$, and such that $|f(n)|\leq \tau_k(n)$ for some $k\in\mathbb{N}$, where τ_k is the k-th divisor function. For $j\geq 0$, let $\widetilde{c_j}$ be the Taylor coefficients about 1 of the function $(s-1)^\kappa F(s)/s$. Then, for any $J\in\mathbb{N}$, and x large enough, we have

$$\sum_{n\leq x} f(n)=x\sum_{j=0}^{J}\widetilde{c_j}\frac{(\log x)^{\kappa-j-1}}{\Gamma(\kappa-j)}+O\left(\frac{x}{(\log x)^{J+2-\Re(\kappa)}}\right).$$

Applying Theorem 2.2.3 to $F(s,\psi)$, we conclude that for any $J\geq 0$, one has

$$\sum_{n\leq x} b(n,\psi)=\delta_\psi\sum_{j=0}^{J}\frac{\tilde{a_j}x}{\Gamma(1/2-j)(\log x)^{1/2+j}}+\mathcal{O}\left(\frac{x}{(\log x)^{J+3/2}}\right), \qquad (2.7)$$

[2] D has no square factors, except maybe 4

where $\delta_\psi = 0$, unless $\psi = \psi_0$, in which case $\delta_{\psi_0} = 1$, and $\tilde{a_j}$ is the j^{th} Taylor coefficient of $(s-1)^{1/2}F(s,\psi_0)/s$. Combining (2.5) with (2.7) we find that for each fixed genus $R \in G(D)$ we have, for any $J \geq 0$,

$$\begin{aligned} B_R(x) &:= \sum_{n\leq x} 1_R(n) \\ &= \frac{1}{|G(D)|}\sum_{n\leq x}\sum_{\psi\in\widehat{G(D)}} b(n,\psi)\psi(R)^{-1} \\ &= \frac{1}{|G(D)|}\sum_{\psi\in\widehat{G(D)}} \psi(R)^{-1}\delta_\psi \sum_{j=0}^{J} \frac{\tilde{a_j}x}{\Gamma(1/2-j)(\log x)^{1/2+j}} + \mathcal{O}\left(\frac{x}{(\log x)^{J+3/2}}\right) \\ &= \frac{1}{|G(D)|}\sum_{j=0}^{J} \frac{\tilde{a_j}x}{\Gamma(1/2-j)(\log x)^{1/2+j}} + \mathcal{O}\left(\frac{x}{(\log x)^{J+3/2}}\right). \end{aligned} \tag{2.8}$$

We then adopt the notation

$$a_j = \frac{\tilde{a}_j}{\Gamma(1/2-j)|G(D)|}. \tag{2.9}$$

We can repeat this argument with the addition of a congruence condition modulo q inserted. Define

$$F(s,\psi,\chi) := \sum_{n\geq 1} b(n,\psi)\chi(n)n^{-s},$$

where χ is a Dirichlet character modulo q. Applying (2.5) as well as the orthogonality relations for Dirichlet characters, one has

$$\frac{1}{\phi(q)|G(D)|}\sum_{\chi \bmod q}\sum_{\psi\in\widehat{G(D)}} b(n,\psi)\chi(n)\chi^{-1}(a)\psi^{-1}(R) = \begin{cases} 1, & 1_R(n) = 1 \textit{ and } n \equiv a \bmod q \\ 0, & \textit{otherwise.} \end{cases} \tag{2.10}$$

Since we are assuming $(q,D) = 1$, then the conductor of χ will be coprime to D. As such, the only choice of characters that makes $F(s,\psi,\chi)$ have a singularity at $s = 1$ is when $\psi = \psi_0$, and $\chi = \chi_0$, the principal character modulo q. We then have

$$\begin{aligned} F(s,\psi_0,\chi_0) &= C_q(s)\left(\zeta(s)A(s)L_D(s)\right)^{1/2} \\ &= C_q(s)F(s,\psi_0), \end{aligned} \tag{2.11}$$

where

$$C_q(s) := \prod_{p|q}(1-p^{-s}) \prod_{\substack{p|q \\ (D|p)=-1}} (1+p^{-s}).$$

Applying Theorem 2.2.3 again, this time to $F(s,\psi_0,\chi_0)$, we conclude that for any $J \geq 0$, one has

$$\sum_{n\le x} b(n,\psi)\chi(n) = \delta_{\psi,\chi}\sum_{j=0}^{J} \frac{\tilde{b_j}x}{\Gamma(1/2-j)(\log x)^{1/2+j}} + \mathcal{O}\left(\frac{x}{(\log x)^{J+3/2}}\right), \tag{2.12}$$

where $\delta_{\psi,\chi}=0$, unless $\psi=\psi_0$ and $\chi=\chi_0$, in which case $\delta_{\psi_0,\chi_0}=1$. Above, $\tilde{b_j}$ is the j^{th} Taylor coefficient of $(s-1)^{1/2}F(s,\psi_0,\chi_0)/s$. Combining (2.10) with (2.12), then for every a such that $(a,q)=1$, one has

$$\sum_{\substack{n\le x\\ n\equiv a \bmod q}} 1_R(n) = \frac{1}{\phi(q)}\frac{1}{|G(D)|}\sum_{j=0}^{J}\frac{\tilde{b_j}x}{\Gamma(1/2-j)(\log x)^{1/2+j}} + \mathcal{O}\left(\frac{x}{(\log x)^{J+3/2}}\right).$$

This proves Theorem 2.2.1, as we see that

$$b_j = \frac{\tilde{b_j}}{\Gamma(1/2-j)\phi(q)|G(D)|},$$

and the constants b_j are independent of R. We can easily express the b_j in terms of the a_j as in the statement of the lemma by comparing the Taylor coefficients of $F(s,\psi_0)$ and $F(s,\psi_0,\chi_0)$; these generating series differ only by the factor $C_q(s)$, as can be seen in (2.11). □

This also concludes the proof of Theorem 2.1.1 by our remarks at the start of this section. Additionally, Theorem 2.2.1 immediately implies Theorem 2.1.2.

Proof of Theorem 2.1.2. We begin by noting that when there is a single form per genus, then $B_f(x;q,a)=B_R(x;q,a)$. As such, Theorem 2.2.1 gives us an explanation for the numerical bias towards the zero residue class; as in (2.4), the constant multiplying the secondary term for $B_f(x;q,a)$ will be

$$\frac{C_q(1)}{\phi(q)|G(D)|}a_1\left(1-\frac{a_0}{a_1}\frac{C_q'(1)}{2C_q(1)}\right)$$

when $a\not\equiv 0 \bmod q$, and

$$\frac{1-C_q(1)}{|G(D)|}a_1\left(1+\frac{a_0}{a_1}\frac{C_q'(1)}{2(1-C_q(1))}\right)$$

when $a\equiv 0 \bmod q$ (again, assuming q is prime). Simplifying the above and noting that $a_0/a_1>0$ immediately yields Theorem 2.1.2. □

2.3 Exceptional Integers in Arithmetic Progressions

In this section, q denotes a prime modulus, and we assume that $C(D)$ is cyclic of odd order h. We will show how the bias towards the zero residue class in this case arises from computing the secondary term in the asymptotic expansion of $B_f(x;q,a)$.

When there is more than one form in the given genus R, we do not have the equality $B_f(x;q,a) = B_f(x;q,a)$ (except in the sense of Theorem 2.1.3). Since we are interested in the secondary of $B_f(x;q,a)$, Theorem 2.1.3 is ineffective due to the $O(x/(\log x)^{2/3})$ error term.

2.3.1 Proof of Theorem 2.1.4

Due to Theorem 2.1.3, studying the secondary term of $B_f(x;q,a)$ is equivalent to estimating the number of integers which are represented by a form from the genus containing f, but not represented by f itself. We refer to these as the "exceptional integers" for f. We define

$$N_f(x,q,a) := \#\{n \leq x : n \equiv a \bmod q, n \textit{ is exceptional for } f\}.$$

It is clear that one has

$$B_f(x;q,a) = B_R(x;q,a) - N_f(x,q,a). \tag{2.13}$$

Summary of Golubeva's Paper

In Theorem 2.1.2, $N_f(x,q,a)$ was always 0. Theorem 2.1.3 can be viewed as a result giving an upper bound on the size of $N_f(x,q,a)$, and a similar result appears in [Gol96]. Computing asymptotics for these exceptional integers turns out to be tricky, with the only result known to the author being the following of Golubeva:

Theorem 2.3.1 ([Gol01])**.** *Let $C(D)$ be cyclic of odd order h. Let $[f^\star]$ be a generator of $C(D)$. Let p_0 be the smallest prime divisor of h, and let $H \subset C(D)$ be the subgroup generated by $[f^\star]^{p_0}$. Let $N_f(x) := N_f(x,1,1)$. Then (writing $[f]$ to be the class that f lies in),*

if $[f] \in H$, one has

$$N_f(x) = A_1 \frac{x}{(\log x)^{1-1/(2p_0)}} (\log\log x)^{p_0-2}(1+o(1)),$$

and if $[f] \notin H$, then one has

$$N_f(x) = A_2 \frac{x}{(\log x)^{1-1/(2p_0)}} (\log\log x)^{p_0-3}(1+o(1)).$$

where A_1, A_2 are positive constants depending only on f.

The constants A_1 and A_2 are not computed explicitly in Golubeva's paper. See Section 2.3.2 for more comments on the computation of these constants, which turn out to be difficult to do explicitly. Golubeva proves Theorem 2.3.1 through a series of lemmas. Lemmas 2-6 of their paper imply that associated to each class of forms $[f] \in C(D)$ is a finite set of tuples $([f_1],[f_2],\cdots [f_r]), \quad [f_i] \in C(D)$, such that once one discards a sparse set, all squarefree integers n which are exceptional for f and coprime to $2D$ will be of the form

$$n = mp_1 \cdots p_r, \tag{2.14}$$

where each prime divisor of m is represented by a form in H, and p_i is any prime represented by $[f_i]$. One also has that

$$r = \begin{cases} p_0 - 2, [f] \in H \\ p_0 - 3, [f] \notin H. \end{cases}$$

Remark. We will consider two tuples $([f_1], \cdots, [f_r])$ and $([g_1], \cdots, [g_r])$ to be distinct if there is no set of indices i,j such that $[f_i] = [g_j]^{\pm 1}$ for $1 \leq i,j \leq r$. If p_i is any prime represented by $[f_i]$, and q_i is any prime represented by $[g_i]$, then $p_1 \cdots p_r$ may be equal to $q_1 \cdots q_r$ if and only if the tuples $([f_1], \cdots, [f_r])$ and $([g_1], \cdots, [g_r])$ are **not** disctinct. This follows from the fact that a prime is uniquely represented by a class of forms and that class' inverse.

There is a more explicit description of these tuples in both cases. Again taking $[f^\star]$ to be a generator of $C(D)$, we may write $[f_i] = [f^\star]^{e_i}$ for each i. Lemmas 5 and 6 of Golubeva's paper imply the following two propositions:

Proposition 1. Let $[f] \in H$ and let p_0 be defined as above. The squarefree integers coprime to $2D$ which are exceptional for f (after discarding a sparse set) are those of the form

$$mp_1p_2 \cdots p_{p_0-2},$$

where each prime divisor of m is represented by a form in H, and p_i is represented by the class of forms $[f_i] = [f^\star]^{e_i}$. The set of choices for $(e_1,e_2,\cdots,e_{p_0-2})$ is the set of diagonal nonzero tuples $(a,a,\cdots,a),\ a \in (\mathbb{Z}/p_0\mathbb{Z})^*$.

Proposition 2. Let $[f] \notin H$ and let p_0 be defined as above. Write $[f] = [f^\star]^{e^\star}$. The squarefree integers coprime to $2D$ which are exceptional for f (after discarding a sparse set) are those of the form

$$mp_1p_2 \cdots p_{p_0-3},$$

Where each prime divisor of m is represented by a form in H, and p_i is represented by the class of forms $[f_i] = [f^\star]^{e_i}$. The set of choices for $(e_1,e_2,\cdots,e_{p_0-3})$ being the set of tuples $(a,a,\cdots,a)$, $a \in \mathbb{Z}/p_0\mathbb{Z})^*$, with the additional condition $a \not\equiv e^\star \bmod p_0$.

Theorem 2.3.1 then follows from a lemma on counting integers of the form (2.14) for each individual tuple $([f_1],\cdots,[f_r])$ which is suitable, and then summing the results over all the distinct tuples:

Lemma 2.3.2. *Let $C(D)$, be cyclic of odd order h. Let $H \subset C(D)$ be a proper subgroup. Fix $[f_1], \dots, [f_r]$, (not necessarily distinct) classes of forms not belonging to H. We wish to count integers of the form*

$$\begin{aligned} &n = mp_1 \cdots p_r, \\ &p|m \implies p \text{ represented by a form } f \in H, \\ &p_j \text{ represented by the class } [f_j]. \end{aligned} \tag{2.15}$$

Let us define $S_{[f_1],\cdots[f_r]}(x,1,1) = S(x)$ by

$$S(x) := \#\{n \leq x : n \text{ squarefree}, (n,2D) = 1, n \text{ satisfying (2.15)}\}.$$

Then we have, as $x \to \infty$,

$$S(x) = A(f_1, \cdots, f_r) \frac{x}{(\log x)^{1-|H|/(2h)}} (\log \log x)^r (1 + o(1)), \tag{2.16}$$

where $A(f_1, \cdots, f_r)$ is a constant depending only on the tuple $[f_1], \cdots, [f_r]$.

As stated in Propositions 1 and 2, there is a finite set of tuples $(f_1, \cdots, f_r)$ for which an integer n of the form of (2.15) can be exceptional for f. Summing (2.16) over all such *distinct*[3] tuples yields Theorem 2.3.1 for squarefree integers coprime to $2D$. The general theorem follows by noting that the inclusion of all integers only changes the asymptotic by a constant factor. See Section 2.3.2 for a comment on the removal of the "squarefree" condition.

Modified Golubeva Lemma and it's Proof

We consider now the behaviour of the exceptional integers in arithmetic progressions, and we obtain the following generalization of Theorem 2.3.2:

Lemma 2.3.3. *Let q be a chosen modulus, $(q,2D) = 1$. Let H, $[f_1], \cdots, [f_r]$ be as in Theorem 2.3.2. Let us define $S_{[f_1],\cdots[f_r]}(x,q,a) = S(x,q,a)$ by*

$$S(x,q,a) := \#\{n \leq x : n \text{ squarefree}, (n,2D) = 1, n \equiv a \bmod q, n \text{ satisfying (2.15)}\}.$$

Then, for $(a,q) = 1$, one has, as $x \to \infty$,

$$S(x,q,a) = \frac{1}{\phi(q)} \prod_{p|q} \left(1 + \frac{\nu_H(p)}{p}\right)^{-1} S(x)(1 + o(1)), \tag{2.17}$$

[3]In the sense of the remark preceding Propositions 1 and 2.

where

$$\nu_H(n) := \begin{cases} 1, n \text{ squarefree, } (n{,}2D) = 1, p|n \implies p \text{ represented by a form } f \in H, \\ 0, \text{ otherwise.} \end{cases}$$

To prove (2.17), we will need to know about the distribution of primes represented by a given quadratic form in an arithmetic progression. The information we need is contained in the following lemma.

Lemma 2.3.4. *Let q be a chosen modulus, $(q{,}2D) = 1$. Fix a binary quadratic form f of discriminant D. We have*

$$\sum_{\substack{p \leq x \\ p \equiv a \bmod q}} 1_f(p) = \frac{1}{\phi(q)} \frac{\delta(f)}{h} Li(x) + O(x \exp(-c\sqrt{\log x})), \tag{2.18}$$

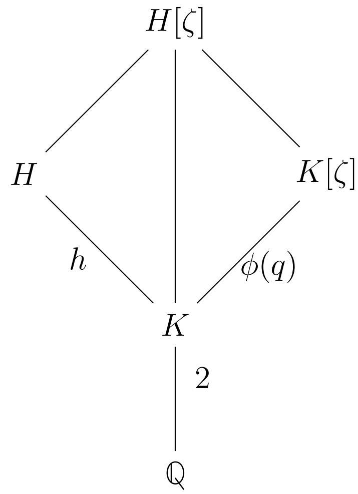
for some positive constant c, where $1_f(n) = 1$ if n is represented by f and $1_f(n) = 0$ otherwise, and

$$\delta(f) = \begin{cases} \frac{1}{2}, & \text{if the class containing } f \text{ has order } \leq 2 \text{ in } C(D) \\ 1, & \text{otherwise.} \end{cases}$$

Proof of Theorem 2.3.4. Let K be the imaginary quadratic field of discriminant D, and let H be the Hilbert class field of K, so that $Gal(H/K) \cong C(D)$. For a given class $\mathcal{C} \in C(D)$, we wish to find an asymptotic for the size of the set

$$S_{\mathcal{C},a} := \{\mathfrak{p} \in \mathcal{O}_K : N(\mathfrak{p}) \leq x, N(\mathfrak{p}) \equiv a \bmod q, \mathfrak{p} \in \mathcal{C}\}.$$

The prime ideals in the set $S_{\mathcal{C},a}$ satisfy two Chebatorev conditions simultaneously, and we may apply the Chebatorev density theorem to count them. First, let $\zeta_q = \zeta = e^{2\pi i/q}$, and consider the following lattice of extensions:



Note that as long as $h > 1$,[4] then $Gal(K[\zeta]/K) \cong (\mathbb{Z}/q\mathbb{Z})^*$, since $\zeta^b \notin K$ for any integer b satisfying $(b,q) = 1$ as K does not contain any q^{th} roots of unity other than ± 1. Since H is the maximal unramified abelian extension of K, and $K[\zeta]/K$ is a finite Galois extension where q is completely ramified, then $H \cap K[\zeta] = K$, and as such we have an isomorphism

$$\begin{aligned} Gal(H[\zeta]/K) &\cong Gal(H/K) \times Gal(K[\zeta]/K) \\ \sigma &\to (\sigma|_H, \sigma|_{K[\zeta]}) \\ Art(\mathfrak{p}, H[\zeta]/K) &\to (Art(\mathfrak{p}, H/K), Art(\mathfrak{p}, K[\zeta]/K)), \end{aligned} \tag{2.19}$$

where Art denotes the Artin symbol. Applying an effective Chebatorev density theorem to the extension $H[\zeta]/K$ tells us that for any given $\sigma \in Gal(H[\zeta]/K)$, one has

$$|\{\mathfrak{p} \in \mathcal{O}_K : N(\mathfrak{p}) \leq x, Art(\mathfrak{p}, H[\zeta]/K) = \sigma\}| = \frac{1}{h\phi(q)} Li(x) + O(x \exp(-c\sqrt{\log x})).$$

Note that the ramified primes contribute only $O(1)$ to our count, and are included in the error term. On the other hand, by the isomorphism of (2.19), for any given $(\mathcal{C}, a) \in C(D) \times (\mathbb{Z}/q\mathbb{Z})^* \cong Gal(H/K) \times Gal(K[\zeta]/K)$, we can find a $\sigma \in Gal(H[\zeta]/K)$ which is mapped to this pair. With this choice of σ in the above equation, one gets

$$\begin{aligned} |S_{\mathcal{C},a}| &= |\{\mathfrak{p} \in \mathcal{O}_K : N(\mathfrak{p}) \leq x, Art(\mathfrak{p}, H/K) = \mathcal{C}, Art(\mathfrak{p}, K[\zeta]/K) = a\}| \\ &= \frac{1}{h\phi(q)} Li(x) + O(x \exp(-c\sqrt{\log x})). \end{aligned}$$

Above, c is some positive constant which can be computed. In fact, in our case one can take $c = (99 * \sqrt{h * \phi(q)})^{-1}$ (see, for example, [Win13], Théorème 1.1 for details on the

[4] In fact, the claim holds as long as $K \neq \mathbb{Q}(\sqrt{m})$ for $m \in \{-1, -3\}$. Either way, we have already dealt with the $h = 1$ case in Theorem 1.

computation of such a constant). To finish the proof, define the sets

$$S'_{\mathcal{C},a} = \{\mathfrak{p} \in \mathcal{O}_K : \mathfrak{p}\overline{\mathfrak{p}} = p, p \equiv a \bmod q, \mathfrak{p} \in \mathcal{C}\},$$

$$S''_{\mathcal{C},a} = \{p \in \mathbb{Q}, p \equiv a \bmod q, p = N(\mathfrak{p}), \mathfrak{p} \in \mathcal{C}\}.$$

We note that

$$|S'_{\mathcal{C},a}| = |S_{\mathcal{C},a}| + O(x^{1/2}),$$

$$|S''_{\mathcal{C},a}| = \sum_{\substack{p \le x \\ p \equiv a \bmod q}} 1_f(n) + O(x^{1/2}), \ \textit{for any } f \in \mathcal{C}.$$

The map $\mathfrak{p} \to N(\mathfrak{p})$ induces a correspondence between $S'_{\mathcal{C},a}$ and $S''_{\mathcal{C},a}$ that is two-to-one if $\mathcal{C} = \mathcal{C}^{-1}$ in $C(D)$. Otherwise, the map is one-to-one. This introduces the factor $\delta(f)$ into our final count for the size of $S''_{\mathcal{C},a}$. □

Proof of Theorem 2.3.3. The function ν_H is multiplicative by definition. We claim that for $(q,2D) = 1$, one has

$$\sum_{\substack{n \le x \\ n \equiv a \bmod q}} \nu_H(n) = \frac{1}{\phi(q)} \sum_{\substack{n \le x \\ (n,q)=1}} \nu_H(n) + O\left(\frac{x}{(\log x)^A}\right) \tag{2.20}$$

for all $A > 0$ i.e. only the principal character mod q will make a contribution to the main term. Indeed, for χ a non-principal character modulo q, one has for $\Re(s) > 1/2$ that

$$\begin{aligned} \sum_{n \ge 1} \frac{\nu_H(n)\chi(n)}{n^s} &= \prod_p \left(1 + \frac{\nu_H(p)\chi(p)}{p^s}\right) \\ &= \exp\left(\sum_p \log\left(1 + \frac{\nu_H(p)\chi(p)}{p^s}\right)\right) \\ &\ll \exp\left(\sum_p \frac{\nu_H(p)\chi(p)}{p^s}\right). \end{aligned} \tag{2.21}$$

By applying Theorem 2.3.4, summing (2.18) over all classes of forms in H, we see that

$$\begin{aligned}\sum_{p\le y}\nu_H(p)\chi(p) &= \sum_{(a,q)=1}\chi(a)\left(\sum_{\substack{p\le y\\ p\equiv a \bmod q}}\nu_H(p)\right)\\
&= \sum_{(a,q)=1}\chi(a)\left(\sum_{\substack{p\le y\\ p\equiv a \bmod q}}1_{f_0}(p)+\frac{1}{2}\sum_{f\in H\backslash[f_0]}\sum_{\substack{p\le y\\ p\equiv a \bmod q}}1_f(p)\right)\\
&= \sum_{(a,q)=1}\chi(a)\left(\frac{1}{2\phi(q)h}Li(y)+\frac{|H|-1}{2\phi(q)h}Li(y)+O_a(y\exp(-c\sqrt{\log y}))\right)\\
&= \frac{|H|}{2\phi(q)h}\sum_{(a,q)=1}\chi(a)\left(Li(y)+O_a(y\exp(-c\sqrt{\log y}))\right)\\
&= O(y\exp(-c\sqrt{\log y})).\end{aligned}$$

We introduced a factor 1/2 in the second line above to avoid double-counting; each prime is simultaneously represented by the class $[f]$ and the class $[f]^{-1}$. Since $|C(D)|$ is odd, $[f]=[f]^{-1}$ if and only if $[f]=[f_0]$, and so there is no double counting for f_0. Hence, by partial summation, we have

$$\begin{aligned}\sum_{p\le y}\frac{\nu_H(p)\chi(p)}{p^s} &= y^{-s}\sum_{p\le y}\nu_H(p)\chi(p)+s\int_2^y t^{-s-1}\sum_{p\le t}\nu_H(p)\chi(p)\,dt\\
&\ll sy^{1-s}\exp(-c\sqrt{\log y})\end{aligned}$$

We see that the limit as y tends to infinity converges if, say $\Re(s)\ge 1-1/(\log y)^{1/2}$. In this range, using (2.21) we have

$$\sum_{n\ge 1}\frac{\nu_H(n)\chi(n)}{n^s}\ll 1.$$

Using partial summation again, with $\Re(s)=1-1/(\log y)^{1/2}$, we have

$$\begin{aligned}\sum_{n\le x}\nu_H(n)\chi(n) &= x^s\sum_{n\le x}\frac{\nu_H(n)\chi(n)}{n^s}-s\int_2^x t^{s-1}\sum_{n\le t}\frac{\nu_H(n)\chi(n)}{n^s}\,dt\\
&\ll x^{1-\frac{1}{(\log x)^{1/2+\epsilon}}}\\
&\ll \frac{x}{(\log x)^A},\end{aligned}$$

for any $A>0$, as the partial sums in the above equation are bounded when $\Re(s)=1-1/(\log y)^{1/2}$.

By applying the orthogonality relations for Dirichlet characters, we have, for $(a,q) = 1$,

$$\begin{aligned}\sum_{\substack{n\le x\\ n\equiv a \bmod q}} \nu_H(n) &= \frac{1}{\phi(q)} \sum_{\chi \bmod q} \chi^{-1}(a) \sum_{n\le x} \chi(n)\nu_H(n) \\ &= \frac{1}{\phi(q)} \sum_{\substack{n\le x\\ (n,q)=1}} \nu_H(n) + O\left(\frac{x}{(\log x)^A}\right).\end{aligned} \tag{2.22}$$

This proves that (2.20) holds for all moduli q satisfying $(q,2D) = 1$. At this stage, one can apply the classical result of [Wir61] for sums of bounded multiplicative functions:

Theorem 2.3.5 (Wirsing)**.** *Given a non-negative multiplicative function $f(n)$, assume there exists constants α,β with $\beta < 2$ such that $f(p^k) \le \alpha\beta^k$ for each prime p and integer $k \ge 2$. Assume further that as $x \to \infty$, one has*

$$\sum_{p\le x} f(p) \sim \kappa \frac{x}{\log x},$$

where κ is a constant. Under these assumptions, as $x \to \infty$, one has

$$\sum_{n\le x} f(n) \sim \frac{e^{\gamma\kappa}}{\Gamma(\kappa)} \frac{x}{\log x} \prod_{p\le x} \left(\sum_{k\ge 0} \frac{f(p^k)}{p^k}\right),$$

where γ is the Euler-Mascheroni constant.

In Lemma 1 of [Gol01], Wirsing's theorem is applied to $\nu_H(n)$, and one finds that $\kappa = \frac{|H|}{2\phi(q)h}$, and

$$\begin{aligned}\sum_{n\le x} \nu_H(n) &= \frac{e^{\gamma\kappa}}{\Gamma(\kappa)} \frac{x}{\log x} \prod_{p\le x} \left(1 + \frac{\nu_H(p)}{p}\right)(1 + o(1)) \\ &= A_3 \frac{x}{\log x} \exp\left(\sum_{p\le x} \frac{\nu_H(p)}{p}\right)(1 + o(1)) \\ &= A_4 \frac{x}{(\log x)^{1-|H|/(2h)}}(1 + o(1)),\end{aligned} \tag{2.23}$$

for some constants A_3, A_4, which can explicitly be written:

$$A_3 := \frac{e^{\gamma\kappa}}{\Gamma(\kappa)} \exp\left(-\sum_{p} \sum_{k\ge 2} \frac{(-1)^k(\nu_H(p))^k}{kp^k}\right),$$

$$A_4 := A_3 \exp\left(m(f_0) + \frac{1}{2} \sum_{\substack{f\in H\\ f\neq f_0}} m(f) \right),$$

where $m(f)$ is the constant defined by the equation

$$\sum_{p\le x} \frac{1_f(p)}{p} = \frac{\delta(f)}{h} \log\log x + m(f) + o(1).$$

And equivalent definition of A_4 would be

$$A_4 := A_3 \lim_{N\to\infty} \exp\left(\sum_{p\le N} \frac{\nu_H(p)}{p} - \frac{|H|}{2h} \log\log N \right)$$

To continue the proof of Theorem 2.3.3, we apply Wirsing's Theorem to the multiplicative function $\nu_H * \chi_0$, where χ_0 is the principal Dirichlet character modulo q. We find, again, that $\kappa = \frac{|H|}{2\phi(q)h}$. This leads to

$$\begin{aligned} \sum_{\substack{n\le x\\ (n,q)=1}} \nu_H(n) &= A_3 \prod_{p|q} \left(1 + \frac{\nu_H(p)}{p}\right)^{-1} \frac{x}{\log x} \exp\left(\sum_{p\le x} \frac{\nu_H(p)}{p} \right) (1+o(1)) \\ &= \prod_{p|q} \left(1 + \frac{\nu_H(p)}{p}\right)^{-1} \sum_{n\le x} \nu_H(n)(1+o(1)). \end{aligned} \tag{2.24}$$

We remark that the constant A_3 in (2.24) and (2.23) are the very same, since

$$\sum_{p\le x} \nu_H(p) \sim \sum_{\substack{p\le x\\ p\nmid q}} \nu_H(p).$$

As such, we obtained the term $\sum_{n\le x} \nu_H(n)$ in (2.24) by using the first equality in (2.23). We can then write

$$S(x,q,a) = \sum_{\substack{p_1\cdots p_r \le \sqrt{x}\\ p_i \text{ repr. by } [f_i]\\ p_i\neq p_j}} \sum_{\substack{m\le x(p_1\cdots p_r)^{-1}\\ m\equiv a(p_1\cdots p_r)^{-1} \bmod q}} \nu_H(m) + \sum_{m\le\sqrt{x}} \nu_H(m) \sum_{\substack{\sqrt{x}\le p_1\cdots p_r\le x/m\\ p_i \text{ repr. by } [f_i]\\ p_i\neq p_j\\ p_1\cdots p_r\equiv m^{-1}a \bmod q}} 1$$

We note that

$$\sum_{m\le\sqrt{x}}\nu_H(m)\sum_{\substack{\sqrt{x}\le p_1\cdots p_r\le x/m\\ p_i \text{ repr. by } [f_i]\\ p_i\ne p_j\\ p_1\cdots p_r\equiv m^{-1}a \bmod q}}1\ll\sum_{m\le\sqrt{x}}\nu_H(m)\sum_{\substack{\sqrt{x}\le p_1\cdots p_r\le x/m\\ p_i\ne p_j}}1$$
$$\ll\frac{x}{\log x}(\log\log x)^{r-1}\sum_{m\le\sqrt{x}}\frac{\nu_H(m)}{m}$$
$$\ll\frac{x}{(\log x)^{1-|H|/(2h)}}(\log\log x)^{r-1},$$

using summation by parts and Wirsing's theorem in the last line. We also have, using (2.22) and (2.24),

$$\sum_{\substack{p_1\cdots p_r\le\sqrt{x}\\ p_i \text{ repr. by } [f_i]\\ p_i\ne p_j}}\sum_{\substack{m\le x(p_1\cdots p_r)^{-1}\\ m\equiv a(p_1\cdots p_r)^{-1} \bmod q}}\nu_H(m)=\frac{1}{\phi(q)}\prod_{p|q}\left(1+\frac{\nu_H(p)}{p}\right)^{-1}\sum_{\substack{p_1\cdots p_r\le\sqrt{x}\\ p_i \text{ repr. by } [f_i]\\ p_i\ne p_j}}\sum_{m\le x(p_1\cdots p_r)^{-1}}\nu_H(m)(1+o(1)),\tag{2.25}$$

and in [Gol01] it is shown that

$$\sum_{\substack{p_1\cdots p_r\le\sqrt{x}\\ p_i \text{ repr. by } [f_i]\\ p_i\ne p_j}}\sum_{m\le x(p_1\cdots p_r)^{-1}}\nu_H(m)=A(f_1,\cdots,f_r)\frac{X}{(\log x)^{1-H/(2h)}}(\log\log x)^r(1+o(1)).$$

This proves Theorem 2.3.3. □

Having proven the validity of (2.17), we may sum it over the disctinct tuples of $[f_1],\cdots[f_r]$ as discussed prior to Theorem 2.3.3. In doing so, we obtain the following result:

Lemma 2.3.6. *Let $N_f'(x,q,a)$ note the number of integers not exceeding x which are square-free, coprime to $2D$, and exceptional for the form f. One has, for a,q,ν_H as in Theorem 2.3.3:*

$$N_f'(x,q,a)=\frac{1}{\phi(q)}\prod_{p|q}\left(1+\frac{\nu_H(p)}{p}\right)^{-1}N_f'(x,1,1)(1+o(1)).$$

Theorem 2.3.6 is a statement about squarefree integers coprime to $2D$, and we also note that

$$B_f'(x,q,a)=B_R'(x,q,a)-N_f'(x,q,a),\tag{2.26}$$

where the $'$ symbol always indicates a sum over squarefree integers which are coprime to $2D$. By our work in this section, we now have an explicit form for the main term of $N'_f(x,q,a)$. This yields Theorem 2.1.4.

2.3.2 Dropping the Squarefreeness Condition

We have proven an asymptotic count for the number of integers not exceeding x which are squarefree, coprime to $2D$, congruent to a mod q, and exceptional for a given form f. We now comment on the claim made (without proof) by Golubeva that ommiting the condition “squarefree" changes our asymptotic count only by a constant. Let f be a fixed reduced form which we decide upon from the start. Let us write $1_E(n)$ to be the function that is 1 when n is exceptional for f, and 0 otherwise.

Lemma 2.3.7. *Given a squarefree number m for which $1_R(m) = 1$, and any integer s, one has*

$$1_E(ms^2) = 1 \implies 1_E(m) = 1.$$

Proof. We prove the contra-positive statement. Suppose that $1_E(m) = 0$, then m is represented by the form f. Since s^2 is represented by the principal form (among other forms), then ms^2 is also represented by f, i.e. $1_E(ms^2) = 0$. □

The converse statement is not true, as we could have some s such that $1_E(m) = 1$, but $1_E(ms^2) = 0$. For example, consider the quadratic form $f(x,y) = x^2 + xy + 15y^2$. In this case, the integer 5 is exceptional for f, but $5 \cdot 3^2$ is actually represented by f, so not exceptional. Thus, $1_E(5) = 1$, and $1_E(5 \cdot 3^2) = 0$. By combining Theorem 2.3.7 with Theorem 2.3.1, we have

$$\begin{aligned}
A_1 x \frac{(\log\log x)^\beta}{(\log x)^\alpha}(1 + o(1)) = \sum_{n \le x} 1_E(n) &= \sum_{s=1}^{\sqrt{x}} \sum_{\substack{ms^2 \le x \\ m \text{ squarefree}}} 1_E(ms^2) \\
&\le \sum_{s=1}^{\infty} \sum_{\substack{m \le x/s^2 \\ m \text{ squarefree}}} 1_E(m) \\
&= A_1 \sum_{s=1}^{\infty} \frac{x}{s^2} \frac{(\log\log x/s^2)^\beta}{(\log x/s^2)^\alpha}(1 + o(1)) \\
&= A_1 \sum_{s=1}^{\infty} \frac{x}{s^2} \frac{(\log\log x)^\beta}{(\log x)^\alpha}\left(1 + O\left(\frac{\log s}{\log\log x}\right)\right) \\
&\le A_1 \frac{\pi^2}{6} x \frac{(\log\log x)^\beta}{(\log x)^\alpha}(1 + o(1)),
\end{aligned}$$

with A_1, α, β above being some constants coming from Theorem 2.3.1. The error term involving $\log s$ can be determined using series expansions. Indeed, one finds that

$$\frac{(\log\log x/s^2)^\beta}{(\log x/s^2)^\alpha} = \frac{(\log\log x)^\beta}{(\log x)^\alpha}\left(1 + O\left(\frac{\log s}{\log\log x}\right)\right).$$

All in all, we've given an upper bound on the number of exceptional integers for f not exceeding x, and this implies that once we drop the squarefree restriction, we only change our count by a constant. Since we have no characterization of the integers where $1_E(m) = 1$ and $1_E(ms^2) = 0$ simultaneously, we cannot seem to explicitly compute the involved constant.

2.4 Conjecture and Observations

In this section, we will state a conjecture on the behaviour of the integers represented by binary quadratic forms in cases not covered by our theorems. It seems natural to conjecture that the bias we have exhibited here in many cases remains true when we drop any restrictions (i.e. we do not assume anything about the class group, and drop any restrictions on squarefreeness/coprimality with $2D$).

Conjecture 1. Let D be any fundamental discriminant. Let f be any reduced binary quadratic form of discriminant D. Let q be any prime such that $(q,2D) = 1$. A numerical bias towards the zero residue class modulo q exists. More precisely, there is a secondary term in the asymptotic expansion of $B_f(x;q,a)$ which has a constant factor that is larger when $a \equiv 0 \bmod q$, and smaller otherwise.

The above conjecture is evidenced by much numerical data (see, for example, Section 2.5), as well as the fact that there always exists a term of size $x/(\log x)^{3/2}$ in the expansion of $B_f(x;q,a)$ that has a constant satisfying the conditions of the conjecture, together with (2.13). With this in mind, proving the conjecture would likely require one to either show that for a given f $N_f(x,q,a)$ exhibits no bias in terms $\gg x/(\log x)^{3/2}$ (as was the case in Theorem 2.1.2), or to show that $N_f(x,q,a)$ exhibits a bias in the correct direction (i.e. $N_f(x,q,a)$ is *smaller* when $a \equiv 0 \bmod q$, and *bigger* otherwise) (as was the case in Theorem 2.1.4).

There seems to be a couple other interesting phenomena which are tangentially related to **??** 1. For example, it seems that there is always an accentuation of the bias toward the zero residue class when one jumps from only considering squarefree integers to considering all integers represented by a given form. One can see an example of this by comparing Table 2.5 with Table 2.6, or by comparing Table 2.7 with Table 2.8.

2.5 Numerical Data

This section contains numerical examples which demonstrate Theorem 2.1.1, Theorem 2.1.2, and Theorem 2.3.6. Table 2.1 and Table 2.2 display values of $B_f(x;q,a)$

for $f(x,y) = x^2 + xy + y^2$. In this case, there is a single reduced form of the chosen discriminant, making it the simplest case covered by our theorems (i.e. we compare with Theorem 2.1.1).

Table 2.3 and Table 2.4 display values of $B_f(x;q,a)$ for $f(x,y) = x^2 + 5y^2$. In this case, there are two separate reduced forms[5] of discriminant -20 lying in two separate genera. These two tables illustrate that the phenomena of Table 2.1 and Table 2.2 carry over to the situation where there is more than one genus, provided that each genus contains only a single form (Theorem 2.1.2).

Table 2.5 provides an example of the behaviour predicted by Theorem 2.3.6. Table 2.6, along with several other numerical investigations suggest that the bias towards the zero residue class is still present when we count all integers, instead of just squarefree integers coprime to $2D$.

Table 2.7 and Table 2.8, demonstrate that a bias towards the zero class still exists in a case where $\nu_H(q) = 0$.

Tables 2.9 and 2.10 give further evidence for the truth of **??** 1 in cases not covered by our theorems.

q	a	$B_f(10^8,7,a)$	Main Term	Two Terms
	0	2342596	2126610	2305520
	1	2181168		2174480
	2	2181169		
7	3	2181008		
	4	2181101		
	5	2181032		
	6	2181096		

Additional Information
$f(x,y) = x^2 + xy + y^2$,
$D = -3$,
$\left(\frac{D}{q}\right) = 1$,
$C(D) = \textit{trivial}$,
$G(D) = \textit{trivial}$,
$B_f(10^8)/q = 2204167$.

Table 2.1: The distribution of $B_f(x;7,a)$ for $f(x,y) = x^2 + xy + y^2$ compared with the first term and first two terms of Theorem 2.1.2. Notice the equidistribution among all residue classes, with a bias towards zero.

q	a	$B_f(10^8,3,a)$	Main Term	Two Terms
	0	4502885	4156480	4448270
3	1	4276237		4262350
	2	4275772		

Additional Information
$f(x,y) = x^2 + 5y^2$,
$D = -20$,
$\left(\frac{D}{q}\right) = 1$,
$C(D) \cong \mathbb{Z}/2\mathbb{Z}$,
$G(D) \cong \mathbb{Z}/2\mathbb{Z}$,
$B_f(10^8)/q = 4351630$.

Table 2.3: The distribution of $B_f(x;3,a)$ for $f(x,y) = x^2 + 5y^2$ compared with the first term and first two terms of Theorem 2.1.2. Notice the similarities with Table 2.1.

[5]The second reduced form of discriminant -20 is $2x^2 + 2xy + 3y^2$. A table of data for this form would look almost identical to Table 2.3 or Table 2.4, depending on the value of $(-20|q)$, since the constants in Theorem 2.2.1 do not depend on the genus.

q	a	$B_f(10^8,5,a)$	Main Term	Two Terms
	0	685734	595452	666121
	1	3685946	3572710	3696540
5	2	3685770		
	3	3685731		
	4	3685990		

Additional Information
$f(x,y) = x^2 + xy + y^2$,
$D = -3$,
$\left(\frac{D}{q}\right) = -1$,
$C(D) = \textit{trivial}$,
$G(D) = \textit{trivial}$,
$B_f(10^8)/q^2 = 617167$,
$(q+1)B_f(10^8)/q^2 = 3703001$

Table 2.2: The distribution of $B_f(x;5,a)$ for $f(x,y) = x^2 + xy + y^2$ compared with the first term and first two terms of Theorem 2.1.2. Notice the equidistribution among all nonzero residue classes, with a much smaller proportion for the zero residue class.

q	a	$B_f(10^8,11,a)$	Main Term	Two Terms
	0	128016	103053	120629
	1	1292745	1236640	1270480
	2	1292628		
	3	1292788		
	4	1292739		
11	5	1292791		
	6	1292573		
	7	1292545		
	8	1292595		
	9	1292875		
	10	1292599		

Additional Information
$f(x,y) = x^2 + 5y^2$,
$D = -20$,
$\left(\frac{D}{q}\right) = -1$,
$C(D) \cong \mathbb{Z}/2\mathbb{Z}$,
$G(D) \cong \mathbb{Z}/2\mathbb{Z}$,
$B_f(10^8)/q^2 = 107892$.
$(q+1)B_f(10^8)/q^2 = 1294700$.

Table 2.4: The distribution of $B_f(x;11,a)$ for $f(x,y) = x^2 + 5y^2$ compared with the first term and first two terms of Theorem 2.1.2. Notice the similarities with Table 2.2.

q	a	$B'_f(10^8,17,a)$
	0	376649
	1	354287
	2	354196
	3	354373
	4	354313
	5	354509
	6	354363
	7	354453
17	8	354278
	9	354228
	10	354259
	11	354418
	12	354329
	13	354263
	14	354347
	15	354402
	16	354192

Additional Information
$f(x,y) = x^2 + xy + 15y^2$,
$D = -59$,
$\left(\frac{D}{q}\right) = 1$,
$C(D) \cong \mathbb{Z}/3\mathbb{Z}$,
$G(D) = trivial$,
$\nu_H(q) = 1$.

Table 2.5: Values of $B'_f(x,3,a)$, where $f(x,y) = x^2 + xy + 15y^2$. A bias towards the zero class is visible, as predicted by Theorem 2.1.4.

q	a	$B_f(10^8,17,a)$
	0	782426
	1	683226
	2	683405
	3	683040
	4	683379
	5	683240
	6	683199
	7	683179
17	8	683380
	9	683427
	10	683042
	11	683073
	12	683018
	13	683403
	14	683214
	15	683499
	16	683323

Additional Information
$f(x,y) = x^2 + xy + 15y^2$,
$D = -59$,
$\left(\frac{D}{q}\right) = 1$,
$C(D) \cong \mathbb{Z}/3\mathbb{Z}$,
$G(D) = trivial$,
$\nu_H(q) = 1$.

Table 2.6: Values of $B_f(x;17,a)$, where $f(x,y) = x^2+xy+15y^2$. A bias still exists numerically when we drop the squarefreeness restriction. The bias seems to become more pronounced in doing so (compare with the bias in Table 2.5).

q	a	$B'_f(10^8,3,a)$
	0	2418331
3	1	2325663
	2	2326169

Additional Information
$f(x,y) = x^2 + xy + 6y^2$,
$D = -23$,
$\left(\frac{D}{q}\right) = 1$,
$C(D) \cong \mathbb{Z}/3\mathbb{Z}$,
$G(D) = trivial$,
$\nu_H(q) = 0$.

Table 2.7: Values of $B'_f(x,3,a)$, where $f(x,y) = x^2 + xy + 6y^2$.

q	a	$B_f(10^8,3,a)$
	0	6223402
3	1	4240799
	2	4239968

Additional Information
$f(x,y) = x^2 + xy + 6y^2$,
$D = -23$,
$\left(\frac{D}{q}\right) = 1$,
$C(D) \cong \mathbb{Z}/3\mathbb{Z}$,
$G(D) = trivial$,
$\nu_H(q) = 0$.

Table 2.8: Values of $B_f(x;3,a)$, where $f(x,y) = x^2 + xy + 6y^2$. Notice how the bias towards 0 seems to become more pronounced when we drop the squarefree restriction (compare with Table 2.7).

q	a	$B_f(10^8,7,a)$
	0	1745576
	1	1254963
	2	1254939
7	3	1254519
	4	1255006
	5	1254481
	6	1254492

Additional Information
$f(x,y) = x^2 + xy + 22y^2$,
$D = -87$,
$\left(\frac{D}{q}\right) = 1$,
$C(D) \cong \mathbb{Z}/6\mathbb{Z}$,
$G(D) \cong \mathbb{Z}/2\mathbb{Z}$,
$\nu_H(q) = 0$.

Table 2.9: Values of $B_f(x;7,a)$, where $f(x,y) = x^2 + xy + 22y^2$. A bias still exists, though none of our theorems cover this case.

q	a	$B_f(10^8,3,a)$
	0	4246393
3	1	3387811
	2	3387781

Additional Information
$f(x,y) = 3x^2 + xy + 8y^2$,
$D = -95$,
$\left(\frac{D}{q}\right) = 1$,
$C(D) \cong \mathbb{Z}/8\mathbb{Z}$,
$G(D) \cong \mathbb{Z}/2\mathbb{Z}$,
$\nu_H(q) = 1$.

Table 2.10: Values of $B_f(x;3,a)$, where $f(x,y) = 3x^2 + xy + 8y^2$. A bias still exists, though none of our theorems cover this case.

Chapter 3

Lemke Oliver and Soundararajan Bias for Consecutive Sums of Two Squares

Chantal David[1], Lucile Devin[2], Jungbae Nam[1] and Jeremy Schlitt[1]

Abstract In a surprising recent work, Lemke Oliver and Soundararajan noticed how experimental data exhibits erratic distributions for consecutive pairs of primes in arithmetic progressions, and proposed a heuristic model based on the Hardy–Littlewood conjectures containing a large secondary term, which fits the data very well. In this paper, we study consecutive pairs of sums of squares in arithmetic progressions, and develop a similar heuristic model based on the Hardy–Littlewood conjecture for sums of squares, which also explains the biases in the experimental data. In the process, we prove several results related to averages of the Hardy–Littlewood constant in the context of sums of two squares.

3.1 Introduction

We study in this paper the distribution of consecutive sums of two squares in arithmetic progressions. Our work is inspired by a recent paper of Lemke Oliver and Soundararajan [LOS16] who proposed a heuristic model based on the Hardy–Littlewood conjecture for the distribution of consecutive primes in arithmetic progressions.

Roughly speaking, it is expected that numbers described by reasonable multiplicative constraints should be well-distributed, in short intervals and in arithmetic progressions. The case of prime numbers is of course well-studied, and this philosophy was also tested

[1]Department of Mathematics & Statistics, Concordia University, 1455 de Maisonneuve Blvd. West, Montréal, Québec, H3G 1M8, CANADA

[2]Univ. Littoral Côte d'Opale, UR 2597 LMPA, Laboratoire de Mathématiques Pures et Appliquées Joseph Liouville, F-62100 Calais, FRANCE

for numbers expressible as sums of two squares, as well as square-free numbers[3]. Gallagher [Gal76] proved that the distribution of primes of size up to x in intervals of size $\log x$ has a Poisson spacing distribution, assuming some explicit form of the Hardy–Littlewood conjecture. This was generalized to primes in arithmetic progressions by Granville [Gra87] and to sums of two squares by Freiberg, Kurlberg and Rosenzweig [FKR17], for intervals of size $\sqrt{\log x}/K$, which is the correct analogue to Gallagher's result in view of (3.1). For primes in larger intervals, Montgomery and Soundararajan [MS04] showed that the spacings exhibit a normal distribution around the mean, assuming again some explicit form of the Hardy–Littlewood conjecture. We prove in this paper a weaker version of their results (Theorem 3.3.4) for the case of sums of two squares which is needed to study the distribution of successive sums of two squares in arithmetic progressions. We speculate that the full analogue of their results can be obtained for sums of two squares, but we did not pursue it as Theorem 3.3.4 is sufficient for our application. Some unexpected irregularities in the distribution of primes in short intervals were discovered by Maier [Mai85], and it was shown by Balog and Wooley [BW00] that sums of two squares exhibit the same irregularities. Sums of two squares in short intervals were also studied over function fields of a finite field $\mathbb{F}_q$, where many results which are inaccessible over number fields can be proven when the size of the finite field $\mathbb{F}_q$ grows [BSSW16, BBSF18, BSF19, GR21].

We first fix some notations. We denote by

$$\mathbf{E} = \{a^2 + b^2 : a,b \in \mathbb{Z}\} = \{E_n : n \in \mathbb{N}\}$$

the set of sums of two squares (enumerated in increasing order), such that E_n is the nth number that can be written as a sum of two squares. Let $\mathbf{1}_{\mathbf{E}}$ be the indicator function of this set. By a classical result of Landau, one has

$$\sum_{n \le x} \mathbf{1}_{\mathbf{E}}(n) \sim K \frac{x}{\sqrt{\log x}}, \tag{3.1}$$

where K is the constant defined by (3.5). The distribution of sums of two squares in arithmetic progressions exhibits different behavior depending on the modulus q of the progression, and we restrict in this paper to the case where q is a prime number such that $q \equiv 1 \pmod 4$. In that case, the sums of squares are equidistributed in all the residue classes $a \pmod q$, including the class $a \equiv 0 \pmod q$ (see Theorem 3.2.2), but unlike the case of the primes, there is a large secondary term depending on if the residue class $a \equiv 0 \pmod q$ or not (see Theorem 3.2.4).

We consider in this paper the following question, which was studied by Lemke Oliver and Soundararajan for primes [LOS16]. Fix a prime number $q \equiv 1 \pmod 4$, and integers

[3] In the case of square-free numbers, the Hardy–Littlewood conjecture is a theorem [Mir49], and the analogue of [LOS16] has been proved recently by Mennema [Men17].

a,b. What is the distribution of

$$N(x;q,(a,b)) := \#\{E_n \le x \;:\; E_n \equiv a \,(\mathrm{mod}\, q),\; E_{n+1} \equiv b \,(\mathrm{mod}\, q)\} \;?$$

Using a model based on randomness, we expect successive sums of two squares to be well-distributed in arithmetic progressions, and each of the q^2 pairs of classes (a,b) to contain the same proportion (asymptotically) of sums of two squares, with possibly a bias towards the pairs (a,b) where $ab \equiv 0 \,(\mathrm{mod}\, q)$ in view of Theorem 3.2.4. However, the numerical data of Table 3.1 (for $q = 5$ and $x = 10^{12}$) shows a lot of fluctuation, and in particular an unexpected large bias against the classes (a,a) including (0,0). Interestingly, this bias goes in the opposite direction of the bias for sums of squares in arithmetic progressions: there are "more" sums of two squares congruent to $0 \,(\mathrm{mod}\, q)$, but there are "less" consecutive sums of two squares congruent to $(0,0) \,(\mathrm{mod}\, q)$.

a	b	$N(10^{12};5,(a,b))$	a	b	$N(10^{12};5,(a,b))$	a	b	$N(10^{12};5,(a,b))$
0	0	4 108 407 474	2	0	8 049 996 586	4	0	7 155 732 959
	1	7 153 121 164		1	5 516 037 772		1	5 356 545 210
	2	5 604 312 560		2	3 754 593 831		2	7 730 855 281
	3	8 054 714 831		3	6 837 553 372		3	5 497 266 920
	4	5 780 373 060		4	5 350 735 550		4	3 768 530 444
1	0	5 777 315 850	3	0	5 609 476 219			
	1	3 765 205 659		1	7 718 021 263			
	2	6 870 009 299		2	5 549 146 140			
	3	5 354 226 097		3	3 765 159 558			
	4	7 742 174 162		4	6 867 117 598			

Table 3.1: $N(x;q,(a,b))$ for $q = 5$ and $x = 10^{12}$. The average of $N(x;q,(a,b))$ is 5 949 465 154.

Estimates for the consecutive sums of squares (or consecutive primes) in arithmetic progressions is a very difficult question, and few results are known. For consecutive primes in arithmetic progressions, it was conjectured by Chowla that there are infinitely many primes p_n such that $p_{n+i-1} \equiv a \,(\mathrm{mod}\, q)$ for $1 \le i \le r$, for any $(a,q) = 1$ and $r \ge 2$. This was proven by Shiu [Shi00]. Recent progress in sieve theory have led to a new proof of Shiu's result [BFTB13], and Maynard has proven that the number of such primes is $\gg \pi(x)$ [May16]. It would be interesting to see if those recent progresses could be applied to get lower bounds for the number of successive sums of two squares E_n such that $E_{n+i-1} \equiv a \,(\mathrm{mod}\, q)$ for $1 \le i \le r$, for any a and $r \ge 2$, but this question was not addressed yet in the literature.

We propose in this paper a heuristic model predicting an asymptotic for $N(x;q,(a,b))$, based on the heuristic of Lemke Oliver and Soundararajan [LOS16] for the case of primes, and exhibiting a similar bias.

Conjecture 3.1.1. Fix a prime $q \equiv 1 \,(\mathrm{mod}\, 4)$, and $J \ge 1$. Then, for any $a \in \mathbb{N}$, we have

$$N(x;q,(a,a)) = \frac{K}{q^2}\frac{x}{\sqrt{\log x}}\left(1-\frac{\sqrt{2}\phi(q)}{\pi}\frac{\sqrt{\log\log x}}{\sqrt{\log x}}+\frac{1}{\sqrt{\log x}}\sum_{j=1}^{J} C_j(\log\log x)^{\frac{1}{2}-j}\right) + O\left(\frac{x}{\log x(\log\log x)^{J+\frac{1}{2}}}\right),$$

for some explicit constants C_j depending only on q. For $a,b\in\mathbb{N}$ with $a\not\equiv b\,(\mathrm{mod}\,q)$, we have

$$N(x;q,(a,b)) = \frac{K}{q^2}\frac{x}{\sqrt{\log x}}\left(1+\frac{\sqrt{2}}{\pi}\frac{\sqrt{\log\log x}}{\sqrt{\log x}}+\frac{C_{a,b}}{\sqrt{\log x}}-\frac{1}{\phi(q)\sqrt{\log x}}\sum_{j=1}^{J} C_j(\log\log x)^{\frac{1}{2}-j}\right) + O\left(\frac{x}{\log x(\log\log x)^{J+\frac{1}{2}}}\right),$$

with

$$C_{a,b} := \frac{1}{2K}\frac{q}{\phi(q)}\sum_{\chi\neq\chi_0}\overline{\chi}(b-a)C_{q,\chi}, \tag{3.2}$$

where the sum is over the non-principal Dirichlet characters modulo q and $C_{q,\chi}$ is defined in (3.28). The value of C_1 is given in Conjecture 3.4.3.

Our heuristic model leading to Conjecture 3.1.1 follows very closely [LOS16], and as such it is based on the Hardy–Littlewood conjectures for sums of squares, which are stated in Section 3.3. Our exposition for that section, and many of the results used for the properties of the (conjectural) Hardy–Littlewood constants for sums of squares follow from [FKR17]. Fix $k\geq 1$ and $\{d_1,\dots,d_k\}\subseteq\mathbb{Z}$. We denote $\mathfrak{S}(\{d_1,\dots,d_k\})$ the Hardy-Littlewood constants for k-tuples of sums of two squares defined in Section 3.3. As the results of [LOS16], our conjecture follows from an average of the Hardy–Littlewood constants, which is one of the main results of our paper.

Theorem 3.1.2. *Let $q\equiv 1\,(\mathrm{mod}\,4)$ be a prime. For each Dirichlet character $\chi\neq\chi_0\,(\mathrm{mod}\,q)$, let $C_{q,\chi}$ be defined by (3.28). Then, for any $J\geq 1$, and $v\neq 0\,(\mathrm{mod}\,q)$, we have*

$$\sum_{h\geq 1}\mathfrak{S}(\{0,h\})e^{-h/H} = H-\frac{2}{K\pi}\sqrt{\log\, H}+\sum_{j=1}^{J} c(j)\,(\log H)^{1/2-j}+O\left((\log H)^{-1/2-J}\right)$$

$$\sum_{\substack{h\geq 1\\ h\equiv 0\,(\mathrm{mod}\,q)}}\mathfrak{S}(\{0,h\})e^{-h/H} = \frac{H}{q}-\frac{2}{K\pi}\sqrt{\log\, H}+\sum_{j=1}^{J} c_0(j)\,(\log H)^{1/2-j}+O\left((\log H)^{-1/2-J}\right)$$

$$\sum_{\substack{h\geq 1\\ h\equiv v\,(\mathrm{mod}\,q)}}\mathfrak{S}(\{0,h\})e^{-h/H} = \frac{H}{q}+\frac{1}{2K^2\phi(q)}\sum_{\substack{\chi\,(\mathrm{mod}\,q)\\ \chi\neq\chi_0}}\bar{\chi}(v)C_{q,\chi}+\sum_{j=1}^{J} c_1(j)\,(\log H)^{1/2-j}+O\left((\log H)^{-1/2-J}\right).$$

The constants $c(j)$ are absolute while the constants $c_0(j),c_1(j)$ depend only on q, they can *all be explicitly computed, in particular the values for $j = 1$ are given in* (3.35) *and* (3.38). *Moreover they satisfy the relation*

$$c_0(j) + \phi(q)c_1(j) = c(j), \quad j \geq 1. \tag{3.3}$$

By using Theorem 3.3.4, which is the analogue of the work [MS04] for sums of two squares, we need only to compute a weighted average of the constants $\mathfrak{S}(\{0,h\})$ associated to 2-tuples, while [FKR17] compute a more general average of the constants $\mathfrak{S}(\{h_1,\dots,h_k\})$ associated to k-tuples. Since the Hardy–Littlewood constants $\mathfrak{S}(\{0,h\})$ can be described explicitly with a simple formula from the work of Connors and Keating [CK97], this allows us to get a very precise result exhibiting a small secondary term which gives the bias. A similar average of the constants $\mathfrak{S}(\{0,h\})$ was computed by Smilansky [Smi13], and we also use some of his results. Moreover, the techniques developed in this paper yield a more precise form of the averages considered in [Smi13] and [FKR17].

Proposition 3.1.3. *Assume the Generalized Riemann Hypothesis. For $\varepsilon > 0$ and $k \geq 2$, we have*

$$\sum_{\substack{1\leq d_1,\dots,d_k\leq H\\ \text{distinct}}} \mathfrak{S}(\{d_1,\dots,d_k\}) = H^k + \frac{k(k-1)H^{k-1}}{\pi K^2}\int_{1/2+\varepsilon}^{1} \frac{F'(\sigma)H^{\sigma-1} + F(\sigma)H^{\sigma-1}\log H}{|\sigma-1|^{1/2}}\,d\sigma + O_{k,\varepsilon}(H^{k-\frac{3}{2}+\varepsilon}),$$

where $F(s) = \zeta(s-1)M(s-1)\left[(s-1)\zeta(s)\right]^{1/2} s^{-1}$, with $M(s)$ as defined by (3.31).

Finally, the heuristic leading to Conjecture 3.1.1 can be generalized to predict an asymptotic for r successive sums of two squares in arithmetic progressions.

Conjecture 3.1.4. Fix a prime $q \equiv 1 \pmod 4$, $r \geq 2$ and $\mathbf{a} = (a_1,\dots,a_r) \in \mathbb{N}^r$. Let

$$N(x;q,\mathbf{a}) := \#\{E_n \leq x \ : \ E_{n+i-1} \equiv a_i \pmod q\}.$$

We have

$$N(x;q,\mathbf{a}) = \frac{x}{q^r}\frac{K}{\sqrt{\log x}}\left(1 + C_{-1}(\mathbf{a})\frac{(\log\log x)^{\frac{1}{2}}}{(\log x)^{\frac{1}{2}}} + \frac{C_0(\mathbf{a})}{(\log x)^{\frac{1}{2}}} + \frac{C_1(\mathbf{a})}{(\log\log x)^{\frac{1}{2}}(\log x)^{\frac{1}{2}}}\right) + O\Big(x(\log\log x)^{-\frac{3}{2}}(\log x)^{-1}\Big),$$

where

$$
\begin{aligned}
C_{-1}(\mathbf{a}) &= \frac{q\sqrt{2}}{\pi}\sum_{i=1}^{r-1}\Big(\tfrac{1}{q} - \delta(a_{i+1} \equiv a_i)\Big) \\
C_0(\mathbf{a}) &= \sum_{\substack{1\le i\le r-1 \\ a_i \not\equiv a_{i+1} \,(\mathrm{mod}\, q)}} C_{a_i,a_{i+1}} \\
C_1(\mathbf{a}) &= -\frac{qC_1}{\phi(q)}\sum_{i=1}^{r-1}\Big(\tfrac{1}{q} - \delta(a_{i+1} \equiv a_i)\Big) + \frac{q\sqrt{2}}{\sqrt{\pi}}\sum_{k=1}^{r-2}\sum_{i=1}^{r-1-k}\frac{\frac{1}{q} - \delta(a_{i+k+1} \equiv a_i)}{k},
\end{aligned}
$$

the constants $C_{a_i,a_{i+1}}$ are defined by (3.2), and the value of C_1 is given in Conjecture 3.4.3.

The structure of the paper is as follows: we review in Section 3.2 the basic properties of sums of two squares, including the secondary terms for the counting function of sums of two squares in arithmetic progressions, which surprisingly we did not find in the literature. We discuss the Hardy–Littlewood conjectures for sums of two squares in Section 3.3. We present the heuristic model leading to Conjecture 3.1.1 in Section 3.4, following Lemke Oliver and Soundararajan [LOS16]; in particular, we explain how the heuristic reduces Conjecture 3.1.1 to an average of Hardy–Littlewood constants (Theorem 3.1.2), which we prove in Section 3.5 using the Selberg–Delange method.

We prove Theorem 3.3.4, which is an analogue of the main result of Montgomery and Soundararajan [MS04] mentioned above used to justify our heuristic, in Section 3.6. We then use this result in Section 3.7 to prove Proposition 3.1.3 which improves the average results of [FKR17] and [Smi13]. Finally, we explain how to deduce Conjecture 3.1.4 from our heuristic in Section 3.8, and we present numerical data in Section 3.9.

Acknowledgements

We thank the organizers of the MOBIUS ANT who invited us to present a preliminary version of the results contained in this paper in a friendly environment with subsequent discussions that led to several improvements in our exposition. We are particularly grateful to Dimitris Koukoulopoulos for constructive discussions, and for pointing out some relevant material from his book [Kou19], and the post of Lucia on Mathoverflow [(ht13]. We thank the anonymous referee for helpful comments.

Our results and conjectures are supported by several numerical data, that were computed using SageMath [The19], PARI/GP [The22] and Mathematica [Inc]. The authors thank the Centre de Recherches Mathématiques (CRM) in the Université de Montréal for offering their clusters for some of the numerical computations.

The first author was supported by a NSERC Discovery Grant and a FQRNT Team Grant; the second author was supported by the grant KAW 2019.0517 from the Knut and Alice Wallenberg Foundation for a post-doctoral position in Chalmers University of Technology and the University of Gothenburg; the fourth author was supported by a Concordia CUSRA.

3.2 Sums of two squares in arithmetic progressions

By a classical result of Landau [Lan09], we have

$$\sum_{n\leq x} \mathbf{1}_{\mathbf{E}}(n) \sim K\frac{x}{\sqrt{\log x}}, \tag{3.4}$$

where

$$K = \frac{1}{\sqrt{2}} \prod_{p\equiv 3 \pmod 4} (1-p^{-2})^{-\frac{1}{2}} \tag{3.5}$$

is the Landau–Ramanujan constant. We remark that, unlike the prime number theorem and contrarily to a claim of Ramanujan (see e.g. [MC99]), the asymptotic above gives only the main term, and there is no simple integral similar to $\mathrm{li}(x)$ which approximates well the number of sums of two squares up to x. This is caused by the fact that the generating series for sums of two squares has an essential singularity at $s = 1$, its contribution is evaluated by the Selberg–Delange method which gives (3.4). It is possible to iterate the Selberg–Delange method to write, for any $J \geq 1$,

$$\sum_{n\leq x} \mathbf{1}_{\mathbf{E}}(n) = Kx\left(\sum_{j=0}^{J} \frac{c_j}{(\log x)^{1/2+j}}\right) + O\left(x(\log x)^{-3/2-J}\right). \tag{3.6}$$

Explicit values for the constants c_j can be found in the literature for $c_0 = 1$ [Lan09], $c_1 = 0.581948659\cdots$ [Sta28, Sta29, Sha64, CLM23] and up to c_{15} in [EG18]. It is possible to get an expression for the number of sums of two squares smaller than x with a better error term, but one loses the simplicity of the formula above as a sum of descending powers of log. We state this result in the next theorem, that we will prove in Section 3.7. A similar expression for the number of sums of two squares exhibiting square-root cancellation under the GRH can be found in [GR21, Theorem B.1], inspired by the work of [RB02]. Such an expression is also suggested in a note of Tenenbaum [Ten15, page 291].

Theorem 3.2.1. *Let $0 < \varepsilon < 1/2$. There exists a constant $c > 0$ such that*

$$\sum_{n\leq x} \mathbf{1}_{\mathbf{E}}(n) = \frac{1}{\pi}\int_{1/2+\varepsilon}^{1} G(\sigma)\frac{x^{\sigma}}{\sigma|\sigma-1|^{1/2}}\, d\sigma + O\left(x\exp\left(-c\sqrt{\log x}\right)\right),$$

where $G(s) = (\zeta(s)(s-1))^{1/2}L(s,\chi_4)^{1/2}(1-2^{-s})^{-1/2}\prod_{p\equiv 3 \pmod 4}\left(1-p^{-2s}\right)^{-1/2}$ and χ_4 is the non-trivial Dirichlet character modulo 4, so $G(s)$ is an analytic function for $\mathrm{Re}(s) > 1/2+\varepsilon$. If we assume the Riemann Hypothesis for $\zeta(s)$ and $L(s,\chi_4)$, we can *replace the error term by* $O\left(x^{1/2+\varepsilon}\right)$.

Even if it is more precise (see Table 3.2), this formula gives somehow less insight on the behaviour of the secondary terms and we come back to the Selberg–Delange method when separating the sums of two squares into congruence classes.

x	Actual	(3.4)	(3.6)	Theorem 3.2.1	(3.4)	(3.6)	Theorem 3.2.1
10^9	173 229 059	167 877 068	172 591 375	173 226 354	1.0319	1.0037	1.00001562
10^{10}	1 637 624 157	1 592 621 708	1 632 873 166	1 637 616 416	1.0283	1.0029	1.00000473
10^{11}	15 570 512 745	15 185 052 177	15 533 945 443	15 570 488 969	1.0254	1.0024	1.00000153
10^{12}	148 736 628 859	145 385 805 874	148 447 838 016	148 736 563 568	1.0230	1.0019	1.00000044

Table 3.2: Comparison of the experimental data for the number of sums of two squares up to x with the asymptotic of (3.4), the asymptotic of (3.6) with the first two terms and the integral of Theorem 3.2.1. The three rightmost columns are the percentage errors. Notice that the error for the integral approximation of Theorem 3.2.1 agrees with the error term under the Riemann Hypothesis.

Let us now consider the distribution of sums of two squares in arithmetic progressions modulo q. For $a \in \mathbb{N}$, following the notations introduced in Section 3.1, let us denote

$$N(x;q,a) := \#\{E_n \leq x \ : \ E_n \equiv a \pmod q\}.$$

The case $q \equiv 1 \pmod 4$ is a prime is particularly simple, and we restrict to that case. We refer the reader to [Rie65, Satz 1] (see also [BW00, Lemma 2.1]) for the general case.

Theorem 3.2.2. *[Rie65, Satz 1] Let* $q \equiv 1 \pmod 4$ *be a prime. Then, for* $a \in \mathbb{Z}/q\mathbb{Z}$,

$$N(x;q,a) := \sum_{\substack{n \leq x \\ n \equiv a \pmod q}} \mathbf{1}_{\mathbf{E}}(n) \sim \frac{K}{q} \frac{x}{\sqrt{\log x}}.$$

If one compares the above theorem with experimental data for $N(x; q, a)$ as shown in Table 3.3, there is a discrepancy, and the experimental data shows an excess for $a \equiv 0 \pmod q$ compared to the other classes modulo q. This is caused by secondary terms that depend on the class a, which do not seem to appear in the literature, and we compute the first such term in Theorem 3.2.4 below. The proof uses the Selberg–Delange method which evaluates the contribution of essential singularities by using Hankel's formula, replacing Cauchy's residue theorem for this case. We state below the version of the method needed for the proof of Theorem 3.2.4, and we refer the reader to [Ten15, Chapter II.5] and [Kou20, Chapter 13], and to Section 3.5 for more details.

Theorem 3.2.3. *[Kou20, Theorem 13.2] Let* $f(n)$ *be a multiplicative function with generating function* $F(s) = \sum_{n\geq 1} f(n)n^{-s}$. *Suppose there exists* $\kappa \in \mathbb{C}$ *such that for* x

large enough

$$\sum_{p\le x} f(p)\log p = \kappa x + O_A\left(x/(\log x)^A\right),$$

for each fixed $A>0$, and such that $|f(n)| \le \tau_k(n)$ for some $k\in\mathbb{N}$, where τ_k is the k-th divisor function. For $j\ge 0$, let $\widetilde{c_j}$ be the Taylor coefficients about 1 of the function $(s-1)^\kappa F(s)/s$. Then, for any $J\in\mathbb{N}$, and x large enough, we have

$$\sum_{n\le x} f(n) = x\sum_{j=0}^{J}\widetilde{c_j}\frac{(\log x)^{\kappa-j-1}}{\Gamma(\kappa-j)} + O\left(\frac{x}{(\log x)^{J+2-\mathrm{Re}(\kappa)}}\right).$$

Theorem 3.2.4. *Let $q\equiv 1 \pmod 4$ be a prime, and let K and c_1 be as defined above. Then,*

$$\sum_{\substack{n\le x\\ n\equiv a \,(\mathrm{mod}\, q)}} \mathbf{1_E}(n) = \frac{K}{q}x\sum_{j=0}^{J}\frac{c_{j,a}}{(\log x)^{1/2+j}} + O\left(\frac{x}{(\log x)^{J+3/2}}\right),$$

where

$$c_{0,a} = c_0 = 1 \quad \text{and} \quad c_{1,a} := \begin{cases} c_1 + \dfrac{\log q}{2} & \text{if } a\equiv 0 \,(\mathrm{mod}\, q)\\ c_1 - \dfrac{\log q}{2(q-1)} & \text{otherwise.}\end{cases} \qquad (3.7)$$

We refer the reader to Table 3.3 for the comparison between the numerical data and Theorem 3.2.4.

q	a	$N(x;q,a)$	Main term	Main and secondary terms
	0	30 700 929 089	29 077 161 174	30 536 403 581
	1	29 508 931 067		29 477 858 608
5	2	29 508 917 111		
	3	29 508 920 778		
	4	29 508 930 814		

Table 3.3: Comparison of the experimental data for $N(x;q,a)$ and the asymptotic of Theorem 3.2.4 using only the main term, or the main term and the first secondary term for $q=5$ and $x=10^{12}$. The average of $N(x;q,a)$ is $\approx$ 29 747 325 771.

Proof. Let $F(s) := \sum_{n\ge 1}\mathbf{1_E}(n)n^{-s}$ be the generating series for sums of two squares. Using the well-known fact that n is a sum of two squares if and only if $v_p(n)$ is even for all primes $p \equiv 3 \pmod 4$, it is easy to see that for $\mathrm{Re}(s) > 1$,

$$\begin{aligned} F^2(s) &= \prod_{p \not\equiv 3 \pmod 4} \left(1 - \frac{1}{p^s}\right)^{-2} \prod_{p \equiv 3 \pmod 4} \left(1 - \frac{1}{p^{2s}}\right)^{-2} \\ &= \zeta(s) L(s, \chi_4) \left(1 - \frac{1}{2^s}\right)^{-1} \prod_{p \equiv 3(4)} \left(1 - \frac{1}{p^{2s}}\right)^{-1} \end{aligned}$$

where χ_4 is the non-principal Dirichlet character modulo 4. Landau [Lan09] also showed that in a neighborhood of $s = 1$,

$$\frac{F(s)}{s^2} = \sum_{\ell \geq 0} i a_\ell (1-s)^{\ell - 1/2},$$

with $a_0 = K\sqrt{\pi}$ and $a_1 = a_0(2c_1 + 1)$ [Sha64]. Applying Theorem 3.2.3 with $\kappa = 1/2$, we get (3.6), using the values a_0, a_1 to get explicit values for the first two Taylor coefficients of $(s-1)^{1/2} F(s)/s$.

To introduce the congruence condition, we write for $a \not\equiv 0 \pmod q$,

$$N(x; q, a) = \frac{1}{q-1} \sum_{\chi \pmod q} \overline{\chi}(a) \sum_{n \leq x} \chi(n) \mathbf{1_E}(n), \tag{3.8}$$

and we denote the generating function of $f_\chi(n) = \chi(n)\mathbf{1_E}(n)$ by $F_\chi(s) := \sum_{n \geq 1} \chi(n)\mathbf{1_E}(n) n^{-s}$. For χ_0 the principal character modulo q and $\chi \neq \chi_0$, we have for $\mathrm{Re}(s) > 1$,

$$\begin{aligned} F_\chi^2(s) &= L(s,\chi) L(s, \chi_4 \chi) \left(1 - \frac{\chi(2)}{2^s}\right)^{-1} \prod_{p \equiv 3(4)} \left(1 - \frac{\chi^2(p)}{p^{2s}}\right)^{-1} \\ F_{\chi_0}^2(s) &= \left(1 - \frac{1}{q^s}\right)^2 F^2(s). \end{aligned} \tag{3.9}$$

For $\chi \neq \chi_0$, since $F_\chi(s)$ is analytic for $\mathrm{Re}(s) > 1/2$, we have for any $\varepsilon > 0$ that

$$\sum_{n \leq x} \chi(n) \mathbf{1_E}(n) = O\left(x^{1/2+\varepsilon}\right),$$

and the theorem will follow by evaluating $\sum_{n \leq x} \chi_0(n) \mathbf{1_E}(n)$ with the Selberg–Delange method. Let $\widetilde{b_j}$ be the Taylor coefficients of $(s-1)^{1/2} F_{\chi_0}(s)/s$ around $s = 1$, and $\widetilde{c_j}$ are the Taylor coefficients of $(s-1)^{1/2} F(s)/s$ around $s = 1$. From (3.9), it is easy to compute

$$\begin{aligned}\widetilde{b_0} &= (1-q^{-1})\,\widetilde{c_0} = (1-q^{-1})\,K\sqrt{\pi}\\ \widetilde{b_1} &= (1-q^{-1})\,\widetilde{c_1} + \frac{\log q}{q}\,\widetilde{c_0} = K\sqrt{\pi}\left(\frac{\log q}{q} - 2c_1(1-q^{-1})\right).\end{aligned}$$

Applying Theorem 3.2.3 with $\kappa = 1/2$, to estimate the sum $\sum_{n\le x}\chi_0(n)\mathbf{1}_{\mathbf{E}}(n)$, and replacing in (3.8), we get the statement of the theorem when $a \not\equiv 0 \,(\mathrm{mod}\, q)$, with

$$\frac{K}{q}c_{0,a} = \frac{\widetilde{b_0}}{(q-1)\Gamma(1/2)}, \qquad \frac{K}{q}c_{1,a} = \frac{\widetilde{b_1}}{(q-1)\Gamma(-1/2)}.$$

For $a \equiv 0 \,(\mathrm{mod}\, q)$, we use the above and (3.6) to obtain

$$\begin{aligned}N(x;q,0) &= \sum_{n\le x}\mathbf{1}_{\mathbf{E}}(n) - \sum_{a\not\equiv 0 \,(\mathrm{mod}\, q)} N(x;q,a)\\ &= \frac{K}{q}x\left(\frac{1}{(\log x)^{1/2}} + \left(c_1 + \frac{\log q}{2}\right)\frac{1}{(\log x)^{3/2}} + \sum_{j=2}^{J}\frac{c_j-(q-1)c_{j,1}}{(\log x)^{1/2+j}}\right) + O\left(x(\log x)^{-3/2-J}\right),\end{aligned}$$

which completes the proof. □

3.3 Hardy–Littlewood conjectures in arithmetic progressions for sum of two squares

We state in this section the analogue of the Hardy–Littlewood prime k-tuple conjectures for the case of sums of two squares, following [FKR17]. We also state new bounds on the average of the Hardy–Littlewood constant in this context that are useful in our heuristic for Conjecture 3.1.1, but are also interesting in themselves as they are related to the distribution of gaps between sums of two squares.

For $k \ge 1$, let $\mathcal{H} = \{h_1, \dots, h_k\} \subseteq \mathbb{Z}$, and

$$R_k(\mathcal{H};x) := \frac{1}{x}\sum_{n\le x}\mathbf{1}_{\mathbf{E}}(n+h_1)\dots\mathbf{1}_{\mathbf{E}}(n+h_k).$$

In the case $\mathcal{H} = \{0\}$, we have

$$R_1(x) := R_1(\{0\};x) = \frac{1}{x}\sum_{n\le x}\mathbf{1}_{\mathbf{E}}(n) \sim \frac{K}{\sqrt{\log x}}.$$

The philosophy of the Hardy–Littlewood conjecture is that the events $\mathbf{1}_{\mathbf{E}}(n+h_i)$ are "independent", and the probability that $n+h_i$ are simultaneously sums of two squares for $1 \le i \le k$ is the product of the probabilities, which is (ignoring the small differences between $\log n$ or

$\log n + h_i)$

$$\left(\frac{K}{\sqrt{\log n}}\right)^k.$$

Of course, the events are not really independent, so we adjust by considering the probabilities that $n + h_i$ are sums of two squares modulo p versus the probably that k independent integers are sums of two squares modulo p. To do so, for each prime p, we define

$$\delta_{\mathcal{H}}(p) = \lim_{\alpha\to\infty} \frac{\#\{0 \le a < p^\alpha : \forall h \in \mathcal{H}, a + h \equiv \square + \square \pmod{p^\alpha}\}}{p^\alpha}.$$

Since $\delta_{\mathcal{H}}(p) = 1$ for $p \equiv 1 \pmod 4$ (see e.g. [FKR17, Proposition 5.1]), we define the singular series for $\mathcal{H} = \{h_1, \dots, h_k\}$ by

$$\mathfrak{S}(\mathcal{H}) := \prod_{p \not\equiv 1 \pmod 4} \frac{\delta_{\mathcal{H}}(p)}{(\delta_{\{0\}}(p))^k}. \tag{3.10}$$

It is proven in [FKR17] that the limit defining $\delta_{\mathcal{H}}(p)$ exists, and the Euler product converges to a non-zero limit provided that $\delta_{\mathcal{H}}(p) > 0$ for all $p \not\equiv 1 \pmod 4$.

Conjecture 3.3.1. [FKR17, Conjecture 1.1] Fix $k \ge 1$, and $\mathcal{H} = \{h_1, \dots, h_k\} \subseteq \mathbb{Z}$. If $\mathfrak{S}(\mathcal{H}) > 0$, then

$$R_k(\mathcal{H}; x) \sim \mathfrak{S}(\mathcal{H})\,(R_1(x))^k \sim \mathfrak{S}(\mathcal{H}) \left(\frac{K}{\sqrt{\log x}}\right)^k$$

This conjecture is still open, but it is known that $\sum_n \mathbf{1_E}(n + h_1) \dots \mathbf{1_E}(n + h_k)$ is infinite for $k = 2{,}3$ by the work of Hooley [Hoo71, Hoo73].

It is not straightforward to give a simple formula for the singular series $\mathfrak{S}(\mathcal{H})$ for a given set $\mathcal{H}$ (see Section 3.6), except the trivial cases $\mathfrak{S}(\emptyset) = \mathfrak{S}(\{h\}) = 1$. For $\mathcal{H} = \{0, h\}$, Connors and Keating [CK97] computed

$$\mathfrak{S}(\{0, h\}) = \frac{1}{2K^2} W_2(h) \prod_{\substack{p \equiv 3 \pmod 4 \\ p \mid h}} \frac{1 - p^{-v_p(h)-1}}{1 - p^{-1}}, \tag{3.11}$$

where

$$W_2(h) = \begin{cases} 1 \text{ if } 2 \nmid h \\ 2 - 3 \cdot 2^{-v_2(h)} \text{ otherwise,} \end{cases}$$

and v_p is the p-adic valuation.

Notice that it means that $\mathfrak{S}(\mathcal{H}) > 0$ when $k = 2$. This can also be proven for $k = 3$, but for general k, we can find sets $\mathcal{H}$ such that $\mathfrak{S}(\mathcal{H}) = 0$. It is easy to see that $\sum_n \mathbf{1_E}(n + h_1) \dots \mathbf{1_E}(n + h_k)$ is finite when $\mathfrak{S}(\mathcal{H}) = 0$.

We now state a slight generalization of the Hardy–Littlewood conjecture where n is restricted to an arithmetic progression modulo q.

Conjecture 3.3.2. (Hardy–Littlewood for sums of two squares in arithmetic progressions) Fix $k \ge 1$, and $\mathcal{H} = \{h_1, \dots, h_k\} \subseteq \mathbb{Z}$. Let $q \equiv 1 \pmod 4$ be a prime, and $a \in \mathbb{Z}$. If $\mathfrak{S}(\mathcal{H}) > 0$, then

$$\begin{aligned} R_k(\mathcal{H};x,q,a) &:= \frac{1}{x}\sum_{\substack{n\le x\\ n\equiv a \,(\mathrm{mod}\, q)}} \mathbf{1}_{\mathbf{E}}(n+h_1)\dots\mathbf{1}_{\mathbf{E}}(n+h_k) \\ &\sim \frac{\mathfrak{S}(\mathcal{H})}{q}\left(\frac{K}{\sqrt{\log x}}\right)^k. \end{aligned}$$

We remark that unlike the generalized Hardy–Littlewood conjecture of [LOS16], we do not need to adjust the local factors at the prime numbers dividing q in $\mathfrak{S}(\mathcal{H})$ since we fixed q to be prime with $q \equiv 1 \,(\mathrm{mod}\, 4)$, and this prime does not appear in the Euler product (3.10) defining $\mathfrak{S}(\mathcal{H})$.

In Conjectures 3.3.1 and 3.3.2, we used $K/\sqrt{\log n}$ for the probability that n is a sum of two squares. As the secondary term for this probability depends on the residue class modulo q from Theorem 3.2.4, we get more precise results by using this second term to refine the probability in Conjecture 3.3.2. We state that in the conjecture below, and we used it to illustrate the fit with the numerical data in Table 3.4, but not in the rest of the paper while getting in the heuristic model leading to Conjecture 3.1.1 and Conjecture 3.1.4 (as those secondary terms would be smaller than some error terms occurring in the heuristic).

Conjecture 3.3.3. (Refined Hardy–Littlewood in arithmetic progressions) Fix $k \ge 1$, and $\mathcal{H} = \{h_1, \dots, h_k\} \subseteq \mathbb{Z}$. Let $q \equiv 1 \,(\mathrm{mod}\, 4)$ be a prime, and $a \in \mathbb{Z}$. If $\mathfrak{S}(\mathcal{H}) > 0$, then

$$R_k(\mathcal{H};x,q,a) = \frac{\mathfrak{S}(\mathcal{H})}{q}K^k\left(\frac{1}{(\log x)^{k/2}} + \frac{1}{(\log x)^{k/2+1}}\sum_{h\in\mathcal{H}} c_{1,h+a} + O\left(\frac{1}{(\log x)^{k/2+2}}\right)\right),$$

where $c_{1,h}$ is defined by (3.7).

a	h	$xR_k(\mathcal{H};x,q,a)$	Main term	Main and secondary term	Err_1	Err_2
0	1	3 906 419 030	3 619 120 683	3 850 620 130	1.0794	1.0145
1	1	3 751 339 794	3 619 120 683	3 718 867 172	1.0365	1.0087
1	2	1 925 818 092	1 809 560 341	1 859 433 586	1.0642	1.0357
0	5	4 062 607 000	3 619 120 682	3 982 373 088	1.1225	1.0201

Table 3.4: Numerical data versus Conjecture 3.3.3 for $\mathcal{H} = \{0, h\}$, $x = 10^{12}$, $q = 5$. The third column shows the numerical data, the 4-th and 5-th columns show the product of x and the prediction of Conjecture 3.3.3 with the main term, and with the main and first secondary term respectively. The last two columns show their percentage errors, respectively.

Finally, we need an equivalent form of Conjecture 3.3.2, inspired by the work of Montgomery and Soundararajan [MS04] for the case of primes, namely

$$\frac{1}{x}\sum_{\substack{n\le x\\ n\equiv a \,(\mathrm{mod}\, q)}}\prod_{h\in\mathcal{H}}\left(\mathbf{1}_{\mathbf{E}}(n+h) - \frac{K}{\sqrt{\log n}}\right) \sim \frac{\mathfrak{S}_0(\mathcal{H})}{q}\left(\frac{K}{\sqrt{\log x}}\right)^{|\mathcal{H}|}. \tag{3.12}$$

Assuming that Conjecture 3.3.2 holds, we get relations between the constants $\mathfrak{S}_0(\mathcal{H})$ and

$\mathfrak{S}(\mathcal{H})$, and it is easy to see that

$$\begin{aligned} \mathfrak{S}_0(\emptyset) &= 1 \\ \mathfrak{S}_0(\{h\}) &= 0 \\ \mathfrak{S}_0(\{h_1, h_2\}) &= \mathfrak{S}(\{h_1, h_2\}) - 1, \end{aligned}$$

and that for a general set $\mathcal{H}$,

$$\mathfrak{S}_0(\mathcal{H}) = \sum_{\mathcal{T} \subseteq \mathcal{H}} (-1)^{|\mathcal{H} \smallsetminus \mathcal{T}|} \mathfrak{S}(\mathcal{T}). \tag{3.13}$$

Mirroring [MS04], we prove in Section 3.6 the following result, which is critical to justify our heuristic.

Theorem 3.3.4. *Let $\mathfrak{S}_0(\mathcal{H})$ the constants defined by (3.13). Then, for any $k \geq 1$ and $\varepsilon > 0$, we have*

$$\sum_{\substack{\mathcal{H} \subseteq [1,h] \\ |\mathcal{H}| = k}} \mathfrak{S}_0(\mathcal{H}) \ll_{k,\varepsilon} h^{\frac{k}{2}+\varepsilon}.$$

Note that our result is weaker than the result of Montgomery and Soundararajan who computed an asymptotic for the average of Theorem 3.3.4 in the case of primes [MS04, Theorem 2]. We did not pursue that as Theorem 3.3.4 is sufficient for our application. We observe that, similar upper bounds are given by Aryan [Ary15b, Ary15a] in the general context of k-tuples of reduced residues.

3.4 Heuristic for the conjecture

We now develop a heuristic leading to Conjecture 3.1.1 following [LOS16]. Let $q \equiv 1 \pmod 4$ be a prime, and we recall that

$$N(x; q, (a,b)) = \#\{E_n \leq x : E_n \equiv a \pmod q, E_{n+1} \equiv b \pmod q\}.$$

We first write

$$N(x; q, (a,b)) = \sum_{\substack{n \leq x \\ n \equiv a \pmod q}} \sum_{\substack{h > 0 \\ h \equiv b-a \pmod q}} \mathbf{1}_{\mathbf{E}}(n) \mathbf{1}_{\mathbf{E}}(n+h) \prod_{t=1}^{h-1} (1 - \mathbf{1}_{\mathbf{E}}(n+t)). \tag{3.14}$$

We introduce the notation

$$\widetilde{\mathbf{1}}_{\mathbf{E}}(n) = \mathbf{1}_{\mathbf{E}}(n) - \frac{K}{\sqrt{\log n}},$$

and for each fixed h in (3.14), we study the sum

$$S_h := \sum_{\substack{n \leq x \\ n \equiv a \pmod q}} \Big(\frac{K}{\sqrt{\log n}} + \widetilde{\mathbf{1}}_{\mathbf{E}}(n) \Big) \Big(\frac{K}{\sqrt{\log (n+h)}} + \widetilde{\mathbf{1}}_{\mathbf{E}}(n+h) \Big) \prod_{0 < t < h} \Big(1 - \frac{K}{\sqrt{\log (n+t)}} - \widetilde{\mathbf{1}}_{\mathbf{E}}(n+t) \Big).$$

If we ignore the small differences among $\sqrt{\log n}$, $\sqrt{\log(n+h)}$, and $\sqrt{\log(n+t)}$ and we expand out the product, we get

$$S_h = \sum_{\mathcal{A}\subset\{0,h\}} \sum_{\mathcal{T}\subset[1,h-1]} (-1)^{|\mathcal{T}|} \sum_{\substack{n\leq x \\ n\equiv a \,(\mathrm{mod}\, q)}} \left(\frac{K}{\sqrt{\log n}}\right)^{2-|\mathcal{A}|} \prod_{\substack{t\in[1,h-1] \\ t\notin\mathcal{T}}} \left(1-\frac{K}{\sqrt{\log n}}\right) \prod_{t\in\mathcal{A}\cup\mathcal{T}} \widetilde{\mathbf{1}}_{\mathbf{E}}(n+t)$$

Finally, denoting

$$\alpha(n) = 1 - \frac{K}{\sqrt{\log n}},$$

and using (3.12), we conjecture that

$$\begin{aligned} S_h &= \sum_{\mathcal{A}\subset\{0,h\}} \sum_{\mathcal{T}\subset[1,h-1]} (-1)^{|\mathcal{T}|} \sum_{\substack{n\leq x \\ n\equiv a \,(\mathrm{mod}\, q)}} \left(\frac{K}{\sqrt{\log(n)}}\right)^{2-|\mathcal{A}|} \alpha(n)^{h-1-|\mathcal{T}|} \prod_{t\in\mathcal{A}\cup\mathcal{T}} \widetilde{\mathbf{1}}_{\mathbf{E}}(n+t) \\ &\sim \frac{x}{q} \sum_{\mathcal{A}\subset\{0,h\}} \sum_{\mathcal{T}\subset[1,h-1]} (-1)^{|\mathcal{T}|} \mathfrak{S}_0(\mathcal{A}\cup\mathcal{T}) \left(\frac{K}{\sqrt{\log x}}\right)^{2+|\mathcal{T}|} \alpha(x)^{h-1-|\mathcal{T}|}. \end{aligned}$$

We emphasize that this is a heuristic argument: in obtaining this expression for S_h, we have not paid attention to the error terms in (3.12), in particular on the dependency on the size of the sets $\mathcal{A}\cup\mathcal{T}$ and on h.

Summing S_h over all $h \equiv b-a \,(\mathrm{mod}\, q)$, this gives the conjectural estimate

$$N(x;q,(a,b)) \sim \frac{x}{q}\alpha(x)^{-1}\left(\frac{K}{\sqrt{\log x}}\right)^2 \mathcal{D}(a,b;x), \tag{3.15}$$

where

$$\mathcal{D}(a,b;x) = \sum_{\substack{h>0 \\ h\equiv b-a \,(\mathrm{mod}\, q)}} \sum_{\mathcal{A}\subset\{0,h\}} \sum_{\mathcal{T}\subset[1,h-1]} (-1)^{|\mathcal{T}|} \mathfrak{S}_0(\mathcal{A}\cup\mathcal{T}) \left(\frac{K}{\alpha(x)\sqrt{\log x}}\right)^{|\mathcal{T}|} \alpha(x)^h. \tag{3.16}$$

In order to evaluate (3.16), we will use the following notations. Let

$$\alpha(x)^h = \left(1-\frac{K}{\sqrt{\log x}}\right)^h = e^{-h/H} \iff H = -\frac{1}{\log\alpha(x)}, \tag{3.17}$$

which implies that

$$\begin{aligned} H &= \frac{\sqrt{\log x}}{K} - \frac{1}{2} + O\left((\log x)^{-1/2}\right) \\ \log H &= \frac{1}{2}\log\log x - \log K + O\left((\log x)^{-1/2}\right). \end{aligned}$$

3.4.1 Discarding the singular series involving larger sets

We approximate $\mathcal{D}(a,b;x)$ by discarding all the singular series where $\mathcal{A}\cup\mathcal{T}$ has more than 2 elements, which is justified by Theorem 3.3.4. We separate in 3 cases, depending on the possible choices for the set $\mathcal{A}\subseteq\{0,h\}$. We use the notation defined in (3.17) for H, and the bound $\sum_{\substack{h>0\\ h\equiv v \pmod q}} \alpha(x)^h h^\ell \ll_\ell H^{\ell+1}$ for any $\ell\geq 0$, and $v\in\mathbb{Z}$.

If $\mathcal{A}=\emptyset$, then for $k\geq 3$, we deduce from Theorem 3.3.4 that

$$\sum_{\substack{h>0\\ h\equiv b-a \pmod q}} \sum_{\substack{\mathcal{T}\subset[1,h-1]\\ |\mathcal{T}|=k}} \mathfrak{S}_0(\mathcal{T})\Big(\frac{K}{\alpha(x)\sqrt{\log x}}\Big)^k \alpha(x)^h \ll_k \Big(\frac{K}{\alpha(x)\sqrt{\log x}}\Big)^k \sum_{\substack{h>0\\ h\equiv b-a \pmod q}} h^{\frac{k}{2}+\varepsilon}\alpha(x)^h$$
$$\ll_k \Big(\frac{K}{\alpha(x)\sqrt{\log x}}\Big)^k H^{1+\frac{k}{2}+\varepsilon} \ll_k (\log x)^{-\frac{k}{4}+\frac{1}{2}+\varepsilon}.$$

If $\mathcal{A}=\{h\}$ and $|\mathcal{A}\cup\mathcal{T}|\geq 3$, we have for $k\geq 2$

$$\sum_{\substack{h>0\\ h\equiv b-a \pmod q}} \sum_{\substack{\mathcal{T}\subset[1,h-1]\\ |\mathcal{T}|=k}} \mathfrak{S}_0(\mathcal{T}\cup\{h\})\Big(\frac{K}{\alpha(x)\sqrt{\log x}}\Big)^k \alpha(x)^h$$
$$\approx_k \frac{1}{q}\Big(\frac{K}{\alpha(x)\sqrt{\log x}}\Big)^k \sum_{\substack{\mathcal{D}\subset[1,H]\\ |\mathcal{D}|=k+1}} \mathfrak{S}_0(\mathcal{D}) \ll_k \frac{1}{q}\Big(\frac{K}{\alpha(x)\sqrt{\log x}}\Big)^k H^{\frac{k+1}{2}+\varepsilon} \ll_k \frac{1}{q}(\log x)^{-\frac{k}{4}+\frac{1}{4}+\varepsilon},$$

where we are approximating the sum over h and $\mathcal{T}$ of the first line as the sum over all $\mathcal{D}$ of size $k+1$ contained in $[1,H]$, which we then bound by Theorem 3.3.4. We obtain the same bound for $\mathcal{A}=\{0\}$ using the fact that $\mathfrak{S}_0$ is invariant by translation.

Finally, in the case $\mathcal{A}=\{0,h\}$, we introduce an extra average. Since $\mathfrak{S}_0$ is translation invariant, we have

$$\sum_{s\geq 1}\mathfrak{S}_0(\{s,t_1{+}s,\dots,t_k{+}s,h{+}s\})e^{-s/H} = \mathfrak{S}_0(\{0,t_1,\dots,t_k,h\})\sum_{s\geq 1}e^{-s/H} \approx \mathfrak{S}_0(\{0,t_1,\dots,t_k,h\})H,$$

and using this, we get for $k \geq 1$

$$\begin{aligned}
\left(\frac{K}{\alpha(x)\sqrt{\log x}}\right)^k & \sum_{\substack{h>0 \\ h\equiv b-a \,(\mathrm{mod}\, q)}} \sum_{\substack{\mathcal{T}\subset[1,h-1] \\ |\mathcal{T}|=k}} \mathfrak{S}_0(\mathcal{T}\cup\{0,h\})\alpha(x)^h \\
&\approx \frac{1}{qH}\left(\frac{K}{\alpha(x)\sqrt{\log x}}\right)^k \sum_{s\geq 1}\sum_{h\geq 1}\sum_{0<t_1<\cdots<t_k<h} \mathfrak{S}_0(\{s,t_1+s,\ldots,t_k+s,h+s\})e^{-(s+h)/H} \\
&\approx \frac{1}{qH}\left(\frac{K}{\alpha(x)\sqrt{\log x}}\right)^k \sum_{0<s<t_1'<\cdots<t_k'<h'<2H} \mathfrak{S}_0(\{s,t_1',\ldots,t_k',h'\}) \\
&\ll_k \frac{1}{q}\left(\frac{K}{\alpha(x)\sqrt{\log x}}\right)^k (2H)^{-1+\frac{k+2}{2}+\varepsilon} \ll_k \frac{1}{q}(\log x)^{-\frac{k}{4}+\varepsilon}.
\end{aligned}$$

Discarding all the singular series where $\mathcal{A}\cup\mathcal{T}$ has more than 2 elements from (3.16), and working again heuristically by ignoring the dependence on $|\mathcal{A}\cup\mathcal{T}|$ in the error terms, we are led to the model

$$\mathcal{D}(a,b;x) = (\mathcal{D}_0+\mathcal{D}_1+\mathcal{D}_2)(a,b;x) + O_\varepsilon((\log x)^{-\frac{1}{4}+\varepsilon}),$$

where

$$\mathcal{D}_0(a,b;x) = \sum_{\substack{h>0 \\ h\equiv b-a \,(\mathrm{mod}\, q)}} \Big(1+\mathfrak{S}_0(\{0,h\})\Big)\alpha(x)^h$$

$$\mathcal{D}_1(a,b;x) = -\left(\frac{K}{\alpha(x)\sqrt{\log x}}\right) \sum_{\substack{h>0 \\ h\equiv b-a \,(\mathrm{mod}\, q)}} \sum_{t\in[1,h-1]} \Big(\mathfrak{S}_0(\{0,t\})+\mathfrak{S}_0(\{t,h\})\Big)\alpha(x)^h$$

$$\mathcal{D}_2(a,b;x) = \left(\frac{K}{\alpha(x)\sqrt{\log x}}\right)^2 \sum_{\substack{h>0 \\ h\equiv b-a \,(\mathrm{mod}\, q)}} \sum_{1\leq t_1<t_2<h} \Big(\mathfrak{S}_0(\{t_1,t_2\})\Big)\alpha(x)^h.$$

Replacing in (3.15), we then conjecture that up to error term of order $x(\log x)^{-\frac{5}{4}+\varepsilon}$, we have

$$N(x;q,(a,b)) \sim \frac{x}{q}\alpha(x)^{-1}\left(\frac{K}{\sqrt{\log x}}\right)^2 (\mathcal{D}_0+\mathcal{D}_1+\mathcal{D}_2)(a,b;x). \tag{3.18}$$

3.4.2 Evaluation of the sums of singular series involving sets of size 2

In order to evaluate (3.18), we first evaluate the simple exponential sums. We will use the notation

$$f(v;q) := \begin{cases} -\frac{1}{2} & v=0 \\ \frac{q-2v}{2q} & 1\leq v\leq q-1 \end{cases}$$

which gives

$$\begin{aligned}
E(H) &:= \sum_{h>0} e^{-h/H} = H - \frac{1}{2} + O(H^{-1}) = \frac{\sqrt{\log x}}{K} - 1 + O\left((\log x)^{-1/2}\right) \\
E(q,v;H) &:= \sum_{\substack{h>0 \\ h \equiv v \,(\mathrm{mod}\, q)}} e^{-h/H} = \frac{H}{q} + f(v;q) + O(H^{-1}) = \frac{\sqrt{\log x}}{Kq} + f(v;q) - \frac{1}{2q} + O\left((\log x)^{-1/2}\right).
\end{aligned}$$

Let

$$\begin{aligned}
S(q,v;H) &:= \sum_{\substack{h\geq 1 \\ h \equiv v \,(\mathrm{mod}\, q)}} \mathfrak{S}(\{0,h\}) e^{-h/H} \\
S_0(q,v;H) &:= \sum_{\substack{h\geq 1 \\ h \equiv v \,(\mathrm{mod}\, q)}} \mathfrak{S}_0(\{0,h\}) e^{-h/H} = \sum_{\substack{h\geq 1 \\ h \equiv v \,(\mathrm{mod}\, q)}} (\mathfrak{S}(\{0,h\}) - 1) e^{-h/H}.
\end{aligned} \tag{3.19}$$

and

$$\begin{aligned}
S(H) &:= \sum_{h\geq 1} \mathfrak{S}(\{0,h\}) e^{-h/H} = \sum_{v \,(\mathrm{mod}\, q)} S(q,v;H) \\
S_0(H) &:= \sum_{h\geq 1} \mathfrak{S}_0(\{0,h\}) e^{-h/H} = \sum_{v \,(\mathrm{mod}\, q)} S_0(q,v;H).
\end{aligned}$$

We then have

$$\begin{aligned}
S_0(q,v;H) &= S(q,v;H) - \frac{H}{q} - f(v;q) + O(H^{-1}) \\
S_0(H) &= S(H) - H + \frac{1}{2} + O(H^{-1}).
\end{aligned} \tag{3.20}$$

Using Theorem 3.1.2, we evaluate $\mathcal{D}_0(a,b;x)$, $\mathcal{D}_1(a,b;x)$ and $\mathcal{D}_2(a,b;x)$.

Proposition 3.4.1. *Let $q \equiv 1 \,(\mathrm{mod}\, 4)$ be a prime. For $j \geq 1$, let $c(j)$ be the constants from Theorem 3.1.2. Then,*

$$\begin{aligned}
&\mathcal{D}_0(a,b;x) + \mathcal{D}_1(a,b;x) + \mathcal{D}_2(a,b;x) \\
&= S(q,b-a;H) + \frac{2}{qK\pi}(\log H)^{1/2} - \frac{1}{2q} - \frac{1}{q}\sum_{j=1}^{J} c(j)(\log H)^{1/2-j} + O\left((\log H)^{-1/2-J} + \frac{\sqrt{\log H}}{\sqrt{\log x}}\right)
\end{aligned}$$

where we use the change of variables (3.17)*. We remark that the error term* $(\log H)^{-1/2-J}$ *is the largest one, for any value of J.*

Proof. First, notice that $\mathcal{D}_0(a,b;x) = S(q, b-a; H)$. For $\mathcal{D}_1(a,b;x)$, we first compute

$$\sum_{\substack{h\geq 2\\ h\equiv b-a \,(\mathrm{mod}\, q)}} \sum_{1\leq t\leq h-1} \mathfrak{S}_0(\{0,t\})e^{-h/H} = \sum_{t\geq 1}\mathfrak{S}_0(\{0,t\})e^{-t/H} \sum_{\substack{h\geq 1\\ h\equiv b-a-t \,(\mathrm{mod}\, q)}} e^{-h/H} \quad (3.21)$$
$$= \left(\frac{H}{q} + O(1)\right) S_0(H),$$

and

$$-\left(\frac{K}{\alpha(x)\sqrt{\log x}}\right) \sum_{\substack{h>0\\ h\equiv b-a \,(\mathrm{mod}\, q)}} \sum_{1\leq t\leq h-1} \mathfrak{S}_0(\{0,t\})e^{-h/H} = \left(-\frac{1}{q} + O\left(\frac{1}{\sqrt{\log x}}\right)\right) S_0(H).$$

We get a similar estimate for the second sum in $\mathcal{D}_1(a,b;y)$ involving $\mathfrak{S}_0(\{t,h\})$ by making a change of variable to replace it by $\mathfrak{S}_0(\{0,r\})$ with $r = h-t$, which gives

$$\mathcal{D}_1(a,b;x) = \left(-\frac{2}{q} + O\left((\log x)^{-1/2}\right)\right) S_0(H).$$

Similarly, for $\mathcal{D}_2(a,b;x)$, we first compute

$$\sum_{\substack{h\geq 3\\ h\equiv b-a \,(\mathrm{mod}\, q)}} \sum_{1\leq t_1<t_2<h} \mathfrak{S}_0(\{t_1,t_2\})e^{-h/H}$$
$$= \sum_{1\leq t_1<t_2} \mathfrak{S}_0(\{0,t_2-t_1\}) \sum_{\substack{h\equiv b-a \,(\mathrm{mod}\, q)\\ h\geq t_2+1}} e^{-h/H} = \sum_{r\geq 1}\mathfrak{S}_0(\{0,r\}) \sum_{t_2\geq r+1} e^{-t_2/H} \sum_{\substack{h'\geq 1\\ h'\equiv b-a-t_2 \,(\mathrm{mod}\, q)}} e^{-h'/H}$$
$$= \sum_{r\geq 1}\mathfrak{S}_0(\{0,r\})e^{-r/H} \sum_{t_2'\geq 1} e^{-t_2'/H} \sum_{\substack{h'\geq 1\\ h'\equiv b-a-t_2'-r \,(\mathrm{mod}\, q)}} e^{-h'/H} = \left(\frac{H^2}{q} + O(H)\right) S_0(H), \quad (3.22)$$

and replacing in the definition of $\mathcal{D}_2(a,b;x)$, we have

$$\mathcal{D}_2(a,b;x) = \left(\frac{1}{q} + O\left((\log x)^{-1/2}\right)\right) S_0(H).$$

Using Theorem 3.1.2 and (3.20) to evaluate $S_0(H)$, this completes the proof. □

One can be more precise regarding the dependence on the congruence classes by separating the sum over t in (3.21) and the sum over r in (3.22) in congruence classes modulo q. In particular, the following refinement of Proposition 3.4.1 will be used for numerical testing. The proof follows directly from the proof of Proposition 3.4.1, and we omit it.

Proposition 3.4.2. *Let $q \equiv 1 \pmod 4$ be a prime. Then*

$$\begin{aligned}
&\mathcal{D}_0(a,b;x) + \mathcal{D}_1(a,b;x) + \mathcal{D}_2(a,b;x) \\
&= E(q, b-a; H) + S_0(q, b-a; H) - 2\frac{K}{\alpha(x)\sqrt{\log x}} \sum_{c \,(\mathrm{mod}\, q)} S_0(q, b-a-c; H) E(q,c; H) \\
&+ \left(\frac{K}{\alpha(x)\sqrt{\log x}}\right)^2 \sum_{c,d \,(\mathrm{mod}\, q)} S_0(q, b-a-c-d; H) E(q,c; H) E(q,d; H).
\end{aligned}$$

3.4.3 Completing the heuristic

We now deduce Conjecture 3.1.1, by replacing Theorem 3.1.2 and Proposition 3.4.1 in (3.18). If $a \equiv b \pmod q$, we have

$$\begin{aligned}
N(x; q,(a,a)) =& \frac{xK^2}{q\log x}\left(1 + \frac{K}{\sqrt{\log x}} + O\left(\frac{1}{\log x}\right)\right)\left[\frac{\sqrt{\log x}}{Kq} - \frac{2(q-1)}{qK\pi}(\log H)^{1/2} - \frac{1}{q}\right. \\
&\left. + \sum_{j=1}^{J}\left(c_0(j) - \frac{c(j)}{q}\right)(\log H)^{1/2-j} + O\left((\log H)^{-J-1/2}\right)\right] \\
=& \frac{Kx}{q^2\sqrt{\log x}}\left[1 + \frac{1}{\sqrt{\log x}}\sum_{j=0}^{J} b_0(j)\,(\log H)^{1/2-j} + O\left(\frac{1}{\sqrt{\log x}(\log H)^{J+1/2}}\right)\right]
\end{aligned} \tag{3.23}$$

where $b_0(0) = -2(q-1)/\pi$ and $b_0(j) = K(qc_0(j) - c(j))$ for $j \geq 1$.

If $a \not\equiv b \pmod q$, we have

$$\begin{aligned}
N(x; q,(a,b)) =& \frac{xK^2}{q\log x}\left(1 + \frac{K}{\sqrt{\log x}} + O\left(\frac{1}{\log x}\right)\right)\left[\frac{\sqrt{\log x}}{Kq} + \frac{2}{qK\pi}(\log H)^{1/2} - \frac{1}{q}\right. \\
&\left. + \frac{1}{2K^2\phi(q)}\sum_{\substack{\chi \,(\mathrm{mod}\, q)\\ \chi\neq\chi_0}} \chi(v)^{-1}C_{q,\chi} + \sum_{j=1}^{J}\left(c_1(j) - \frac{c(j)}{q}\right)(\log H)^{1/2-j} + O\left((\log H)^{-J-1/2}\right)\right] \\
=& \frac{Kx}{q^2\sqrt{\log x}}\left[1 + C_{a,b} + \frac{1}{\sqrt{\log x}}\sum_{j=0}^{J} b_1(j)\,(\log H)^{1/2-j} + O\left((\log H)^{-J-1/2}\right)\right] \qquad (3.24) \\
=& \frac{Kx}{q^2\sqrt{\log x}}\left[1 + C_{a,b} - \frac{1}{\phi(q)}\frac{1}{\sqrt{\log x}}\sum_{j=0}^{J} b_0(j)\,(\log H)^{1/2-j} + O\left((\log H)^{-J-1/2}\right)\right],
\end{aligned}$$

where $C_{a,b} = \frac{q}{2K\phi(q)}\sum_{\substack{\chi \,(\mathrm{mod}\, q)\\ \chi\neq\chi_0}} \chi(v)^{-1}C_{q,\chi}$, $b_1(0) = 2/\pi$ and $b_1(j) = K(qc_1(j) - c(j))$ for $j \geq 1$.

For the last line, we used (3.3) which gives $b_1(j) = -\frac{b_0(j)}{\phi(q)}$, for $j \geq 0$.

To deduce Conjecture 3.1.1 and obtain the explicit expressions for the constants C_j for $0 \leq j \leq J$, from the above expressions (3.23) and (3.24), we approximate $(\log H)^{1/2-j}$ for $0 \leq j \leq J$, where H is given by (3.17). We illustrate the process below for $J = 1$.

Using the approximations

$$(\log H)^{1/2} = \frac{1}{\sqrt{2}}\sqrt{\log\log x} - \frac{\log K}{\sqrt{2}}\frac{1}{\sqrt{\log\log x}} + O\left((\log\log x)^{-3/2}\right),$$
$$(\log H)^{-1/2} = \frac{\sqrt{2}}{\sqrt{\log\log x}} + O\left((\log\log x)^{-3/2}\right),$$

we obtain

$$\sum_{j=0}^{1} b_0(j)\ (\log H)^{1/2-j} = -\frac{\sqrt{2}(q-1)}{\pi}\sqrt{\log\log x} + \left(\frac{\sqrt{2}(q-1)\log K}{\pi} + \sqrt{2}b_0(1)\right)(\log\log x)^{-1/2} + O\left((\log\log x)^{-3/2}\right) \tag{3.25}$$

Using the values of $c(1)$ and $c_0(1)$ given by (3.35) and (3.38), we have

$$b_0(1) = K(qc_0(1) - c(1)) = \frac{\phi(q)}{\pi}\left(\frac{\omega+\gamma}{2} + \frac{q}{\phi(q)}\log q\right),$$

where γ is the Euler-Mascheroni constant and ω is defined in Lemma 3.5.2. Replacing in (3.25) and then in (3.23), we get Conjecture 3.4.3 below which is the special case of Conjecture 3.1.1 for $J = 1$. The case $a \not\equiv b \pmod q$ follows from multiplying the corresponding term by $\frac{-1}{\phi(q)}$ in (3.24). The general case of Conjecture 3.1.1 follows similarly by using approximations for $(\log H)^{-1/2-j}$ as above for $1 \leq j \leq J$.

Conjecture 3.4.3. Fix $q \equiv 1 \pmod 4$. Then,

$$N(x,q,(a,a)) \sim \frac{x}{q^2}\frac{K}{\sqrt{\log x}}\left(1 - \frac{\sqrt{2}\phi(q)}{\pi}\frac{(\log\log x)^{1/2}}{(\log x)^{1/2}} + \frac{C_1}{(\log x)^{1/2}(\log\log x)^{1/2}}\right)$$

up to an error term of $O\left(\frac{x}{\log x(\log\log x)^{3/2}}\right)$, and with

$$C_1 = \frac{\sqrt{2}\phi(q)}{\pi}\left(\log K + \frac{\omega+\gamma}{2}\right) + \frac{\sqrt{2}q\log q}{\pi},$$

where γ is the Euler-Mascheroni constant and ω is defined in Lemma 3.5.2.

For $a \neq b \pmod q$,

$$N(x,q,(a,b)) = \frac{x}{q^2}\frac{K}{\sqrt{\log x}}\left(1 + \frac{\sqrt{2}}{\pi}\frac{\sqrt{\log\log x}}{\sqrt{\log x}} + \frac{C_{a,b}}{\sqrt{\log x}} - \frac{C_1}{\phi(q)(\log x)^{1/2}(\log\log x)^{1/2}}\right)$$

up to an error term of $O\left(\frac{x}{\log x(\log\log x)^{3/2}}\right)$, and with

$$C_{a,b} := \frac{1}{2K}\frac{q}{\phi(q)}\sum_{\chi\neq\chi_0}\overline{\chi}(b-a)C_{q,\chi}$$

where the sum is over the non-principal Dirichlet characters modulo q and $C_{q,\chi}$ is defined in (3.28).

3.5 Proof of Theorem 3.1.2

Proof. As in [Smi13], we define $a(h) = 2K^2\mathfrak{S}(\{0,h\})$. Then using (3.11), we see that $a(h)$ is a multiplicative function of h with

$$a(p^k) = \begin{cases} 1 & \text{for } p \equiv 1 \pmod 4, \\ 2 - \dfrac{3}{2^k} & \text{for } p = 2, k \geq 1, \\ \dfrac{1 - p^{-(k+1)}}{1 - p^{-1}} & \text{for } p \equiv 3 \pmod 4. \end{cases}$$

Using Mellin Inversion, we have

$$2K^2 S(H) = \sum_{h \geq 1} a(h) e^{-h/H} = \frac{1}{2\pi i} \int_{(2)} D(s) H^s \Gamma(s) ds,$$

where $D(s) = \sum_{h \geq 1} a(h) h^{-s}$. Similarly, for χ a character modulo q,

$$2K^2 S(H, \chi) := \sum_{h \geq 1} a(h)\, \chi(h) e^{-h/H} = \frac{1}{2\pi i} \int_{(2)} D_\chi(s) H^s \Gamma(s) ds, \tag{3.26}$$

where $D_\chi(s) = \sum_{h \geq 1} a(h)\chi(h) h^{-s}$.

In order to compute $S(H)$ and $S(H,\chi)$, we move the contour integral and pick up the contributions of the singularities of the integrand. So, first, we need to understand the analytic properties of the generating series $D(s)$ and $D_\chi(s)$. Using the formulas for $a(p^k)$ above, we have

$$D_\chi(s) = R_\chi(s) P_\chi(s) Q_\chi(s),$$

where

$$\begin{aligned} R_\chi(s) &= 1 + 2\left(\frac{\chi(2)2^{-s}}{1 - \chi(2)2^{-s}}\right) - 3\left(\frac{\chi(2)2^{-(s+1)}}{1 - \chi(2)2^{-(s+1)}}\right) \\ P_\chi(s) &= \prod_{p \equiv 1\ (4)} (1 - \chi(p) p^{-s})^{-1} \\ Q_\chi(s) &= \prod_{p \equiv 3\ (4)} (1 - \chi(p) p^{-s})^{-1} (1 - \chi(p) p^{-(s+1)})^{-1}. \end{aligned}$$

This can be rewritten as

$$D_\chi(s) = L(s,\chi)(1 - \chi(2)2^{-s}) R_\chi(s) Q_{1,\chi}(s) = L(s,\chi) L(s+1,\chi)^{\frac{1}{2}} M_\chi(s)$$

where

$$
\begin{aligned}
Q_{1,\chi}(s) &= \prod_{p\equiv 3\,(4)} (1-\chi(p)p^{-(s+1)})^{-1} \\
&= \Big(\frac{L(s+1,\chi)}{L(s+1,\chi\cdot\chi_4)}\Big)^{\frac{1}{2}}(1-\chi(2)2^{-(s+1)})^{\frac{1}{2}} \prod_{p\equiv 3\,(\mathrm{mod}\,4)} (1-\chi(p)^2p^{-2(s+1)})^{-\frac{1}{2}}, \\
M_\chi(s) &= \left(1+\chi(2)2^{-s}-3\Big(\frac{\chi(2)2^{-(s+1)}(1-\chi(2)2^{-s})}{1-\chi(2)2^{-(s+1)}}\Big)\right)L(s+1,\chi\cdot\chi_4)^{-\frac{1}{2}} \\
&\quad\times(1-\chi(2)2^{-(s+1)})^{\frac{1}{2}} \prod_{p\equiv 3\,(\mathrm{mod}\,4)} (1-\chi(p)^2p^{-2(s+1)})^{-\frac{1}{2}} \\
&= (1-\chi(2)2^{-s}+\chi(4)2^{-2s})L(s+1,\chi\cdot\chi_4)^{-\frac{1}{2}} \\
&\quad\times(1-\chi(2)2^{-(s+1)})^{-\frac{1}{2}} \prod_{p\equiv 3\,(\mathrm{mod}\,4)} (1-\chi(p)^2p^{-2(s+1)})^{-\frac{1}{2}},
\end{aligned}
$$

where χ_4 is the primitive character modulo 4. The formula for $Q_{1,\chi}(s)$ follows from developing the identity $(1-\chi(p)^2p^{-2(s+1)}) = (1-\chi(p)p^{-(s+1)})(1+\chi(p)p^{-(s+1)})$. Since $\chi \neq \chi_4$, the function M_χ is holomorphic in the half plane $\mathrm{Re}(s) \geq 0$, and we can push this limit a bit further to the left depending on the zero-free region of $L(s+1,\chi\cdot\chi_4)$ (up to $\mathrm{Re}(s) > -\frac{1}{2}$ under Riemann Hypothesis).

In the case χ is a non-principal character, $L(s,\chi)$ is entire on the complex plane, and $L(s+1,\chi)^{\frac{1}{2}}$ is holomorphic in a region containing the half-plane $\mathrm{Re}(s) \geq 0$, where $L(s+1,\chi)$ does not vanish. As such, there is no pole or singularity in the integrand at $s = 1$. If we shift the line of integration of (3.26) to the left of the line $\mathrm{Re}(s) = 0$ using the standard zero-free region and estimates for L-functions, we obtain (for some $c > 0$ and any $\varepsilon > 0$)

$$
S(H,\chi) = \frac{C_{q,\chi}}{2K^2} + \begin{cases} O(H^{-1/2+\varepsilon}) & \text{under GRH} \\ O\left(\exp\left(-c\sqrt{\log H}\right)\right) & \text{otherwise,} \end{cases} \tag{3.27}
$$

where the constant term comes from the contribution of pole of order 1 from $\Gamma(s)$ at $s = 0$, and

$$
\begin{aligned}
C_{q,\chi} &= D_\chi(0) = L(0,\chi)L(1,\chi)^{\frac{1}{2}}M_\chi(0) \qquad (3.28)\\
&= L(0,\chi)L(1,\chi)^{\frac{1}{2}}L(1,\chi\cdot\chi_4)^{-\frac{1}{2}}(1-\chi(2)+\chi(4))(1-\chi(2)2^{-1})^{-\frac{1}{2}} \prod_{p\equiv 3\,(\mathrm{mod}\,4)} (1-\chi(p)^2p^{-2})^{-\frac{1}{2}}.
\end{aligned}
$$

Note that $C_{q,\chi} \neq 0$ only when $\chi(-1) = -1$.

If $\chi = \chi_0$, $D_{\chi_0}(s)$ has a simple pole at $s = 1$ with residue $2K^2\phi(q)/q$, and no other singularities

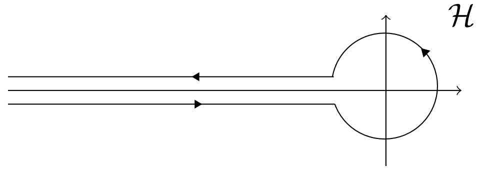


Figure 3.1: Hankel's Contour

for $\mathrm{Re}(s) > 0$, and we move the integral to $\mathrm{Re}(s) = \varepsilon > 0$. This gives

$$\begin{aligned}\sum_{h\geq 1} a(h)\chi_0(q)e^{-h/H} &= \frac{1}{2\pi i}\int_{(2)} D_{\chi_0}(s)H^s\Gamma(s)ds \\ &= 2\frac{\phi(q)}{q}K^2H + \frac{1}{2\pi i}\int_{(\varepsilon)} D_{\chi_0}(s)H^s\Gamma(s)ds. \end{aligned} \tag{3.29}$$

Similarly, we have that

$$\sum_{h\geq 1} a(h)e^{-h/H} = \frac{1}{2\pi i}\int_{(2)} D(s)H^s\Gamma(s)ds = 2K^2H + \frac{1}{2\pi i}\int_{(\varepsilon)} D(s)H^s\Gamma(s)ds, \tag{3.30}$$

where

$$D(s) = \zeta(s)(1-2^{-s})R(s)Q_1(s) = \zeta(s)\zeta(s+1)^{\frac{1}{2}}M(s), \tag{3.31}$$

and the functions R, Q_1, M are obtained by taking $\chi \equiv 1$ in the previous definitions.

To account for the contribution of the singularity of $D_{\chi_0}(s)$ and $D(s)$ as $s = 0$ to the integrals (3.29) and (3.30), we use again the Selberg–Delange method. Since we are evaluating a Mellin transform, we cannot use directly [Kou20, Theorem 13.2] as in Section 3.2, but we are following the same standard steps. We first approximate the line of integration $\Re(s) = \varepsilon$ by the truncated segment from $\varepsilon - iT$ to $\varepsilon + iT$, which we then deform to a truncated Hankel's contour. This is possible since there are no residue inside this contour. We then replace this contour by the infinite Hankel's countour $\mathcal{H}$ of Figure 3.1 with a very good error term, which allows us to use Theorem 3.5.1 to compute the contribution of the singularity of the generating functions (for each term of the Taylor series). We refer the reader to [Kou20, Chapter 13] and [Ten15, Chapter 5] for more details. The contributions to $S(H)$ and $S(H, \chi_0)$ will be different in magnitude, because the singularities of $\zeta(s)\zeta(s+1)^{1/2}$ and $L(s,\chi_0)L(s+1,\chi_0)^{1/2}$ at $s = 0$ are different, since $L(0, \chi_0) = 0$, but $\zeta(0) \neq 0$.

Theorem 3.5.1 (Hankel's formula [Ten15] Theorem 0.17 p.179)**.** *Fix any $r > 0$, and let $\mathcal{H}$ be the Hankel's countour, which is the path consisting of the circle $|s| = r$ excluding the point $s = -r$, and of the half-line $(-\infty, -r]$ covered twice, with respective arguments π and $-\pi$. Then for any complex number z, we have*

$$\frac{1}{2\pi i}\int_{\mathcal{H}} s^{-z}e^s\, ds = \frac{1}{\Gamma(z)}.$$

We first work with $D(s) = \zeta(s)\zeta(s+1)^{\frac{1}{2}}M(s)$. The function $M(s)$ is analytic around $s = 0$, and

$$M(0) = 2KL(1, \chi_4)^{-\frac{1}{2}} = \frac{4K}{\sqrt{\pi}}.$$

Then, $s^{\frac{3}{2}}D(s)\Gamma(s)$ is analytic and non-zero around $s = 0$ with Taylor series $\sum_{n\geq 0} c_n s^n$, and we write

$$D(s)\Gamma(s) = \frac{a(3/2)}{s^{3/2}} + \frac{a(1/2)}{s^{1/2}} + a(-1/2)s^{1/2} + \dots.$$

We now compute the contribution to the integral (3.30) for each term of the series above using Theorem 3.5.1. For every term $a(z)/s^z$ of the Taylor series above (where $z = 3/2, 1/2, -1/2, \dots$), we have

$$\begin{aligned}\frac{1}{2\pi i}\int_{\mathcal{H}} a(z)s^{-z}H^s ds &= \frac{a(z)}{2\pi i}\int_{\mathcal{H}} s^{-z}e^{s\log H}ds \\ &= \frac{a(z)(\log H)^{z-1}}{2\pi i}\int_{\mathcal{H}} t^{-z}e^t dt = \frac{a(z)(\log H)^{z-1}}{\Gamma(z)},\end{aligned}$$

where we used the change of variables $t = s\log H$. This gives, for any integer $N \geq 1$,

$$\frac{1}{2\pi i}\int_{(\varepsilon)} D(s)H^s\Gamma(s)ds = \sum_{n=0}^{N} \frac{a(3/2-n)(\log H)^{1/2-n}}{\Gamma(3/2-n)} + O\left((\log H)^{1/2-N-1}\right).$$

Replacing in (3.30), this gives

$$S(H) = H + \sum_{j=0}^{J} c(j)(\log H)^{1/2-j} + O\left((\log H)^{-1/2-J}\right), \tag{3.32}$$

with

$$c(j) = \frac{a(3/2-j)}{2K^2\Gamma(3/2-j)}, \quad j \geq 0.$$

To complete the proof of Theorem 3.1.2, we now compute the values of $c(0)$ and $c(1)$. Using the expansions[4] around $s = 0$

$$\sqrt{\zeta(s+1)} = \frac{1}{s^{1/2}}\left(1 + \frac{\gamma}{2}s + O(s^2)\right), \qquad \Gamma(s) = \frac{1}{s} - \gamma + O(s),$$

we have

$$D(s)\Gamma(s) = \frac{1}{s^{\frac{3}{2}}}\zeta(s)M(s)\left(1 - \frac{\gamma}{2}s + O(s^2)\right), \tag{3.33}$$

[4]We used the Laurent expansion of $\zeta(s)$ at $s = 1$

$$\zeta(s) = \frac{1}{s-1} + \sum_{n=0}^{\infty}(-1)^n\frac{\gamma_n}{n!}(s-1)^n, \quad \text{with } \gamma_n := \lim_{N\to\infty}\left(\sum_{1\geq k\geq N}\frac{\log^n k}{k} - \int_1^N \frac{\log^n t}{t}dt\right), \quad \text{and } \gamma := \gamma_0.$$

which gives

$$a(3/2) = \zeta(0)M(0) = -\frac{2K}{\sqrt{\pi}} \implies c(0) = \frac{a(3/2)}{2K^2\Gamma(3/2)} = -\frac{2}{K\pi}.$$

To get the value of $c(1)$, we need the first 2 terms of the Taylor series around $s = 0$ of the analytic function

$$Z(s) = \zeta(s)M(s) = Z(0) + Z'(0)s + O\left(s^2\right). \tag{3.34}$$

Replacing in (3.33), and using Lemma 3.5.2 for the value of $Z'(0)$, we have

$$a(1/2) = \left(Z'(0) - \frac{\gamma}{2}Z(0)\right) = \frac{K(\omega+\gamma)}{\sqrt{\pi}} \implies c(1) = \frac{a(1/2)}{2K^2\Gamma(1/2)} = \frac{\omega+\gamma}{2\pi K}, \tag{3.35}$$

where γ is the Euler-Mascheroni constant and ω is defined in Lemma 3.5.2.

We now turn to the secondary term for the sum

$$\begin{aligned}\sum_{h\geq 1} a(h)\chi_0(h)e^{-h/H} &= \frac{1}{2\pi i}\int_{(2)} D_{\chi_0}(s)H^s\Gamma(s)ds \\ &= 2\frac{\phi(q)}{q}K^2H + \frac{1}{2\pi i}\int_{(\varepsilon)} D_{\chi_0}(s)H^s\Gamma(s)ds \end{aligned} \tag{3.36}$$

which is similar to the above replacing $D(s)$ with D_{χ_0}, where χ_0 is the principal character modulo q. We have

$$D_{\chi_0}(s) = L(s,\chi_0)L(s+1,\chi_0)^{\frac{1}{2}}M_{\chi_0}(s),$$

where

$$M_{\chi_0}(s) = (1 - 2^{-s} + 2^{-2s})L(s+1,\chi_0\cdot\chi_4)^{-\frac{1}{2}}(1-2^{-(s+1)})^{-\frac{1}{2}}\prod_{p\equiv 3\,(\mathrm{mod}\,4)}(1-p^{-2(s+1)})^{-\frac{1}{2}}$$

since $\chi_0(p) = 1$ for each $p \nmid q$ and $q \equiv 1\,(\mathrm{mod}\,4)$. We remark that $M_{\chi_0}(s) = (1 - q^{-(s+1)})^{-\frac{1}{2}}M(s)$, which implies that

$$D_{\chi_0}(s) = L(s,\chi_0)\zeta(s+1)^{\frac{1}{2}}M(s) = (1-q^{-s})D(s).$$

Then writing $(1-q^{-s}) = (\log q)s + O(s^2)$, we notice that $s^{1/2}D_{\chi_0}(s)\Gamma(s)$ is analytic and non-zero around $s = 0$. Indeed, $L(s,\chi_0)$ has a simple zero at $s = 0$ which cancels the pole of $\Gamma(s)$. Around $s = 0$, we write

$$D_{\chi_0}(s)\Gamma(s) = \frac{b(1/2)}{s^{1/2}} + b(-1/2)s^{1/2} + b(-3/2)s^{3/2} + \ldots,$$

and working as above this gives

$$\frac{1}{2\pi i}\int_{(\varepsilon)} D_{\chi_0}(s)H^s\Gamma(s)ds = \sum_{n=0}^{N}\frac{b(1/2-n)(\log H)^{-1/2-n}}{\Gamma(1/2-n)} + O\left((\log H)^{-1/2-N-1}\right),$$

replacing in (3.36), we have

$$S(H,\chi_0) = \frac{\phi(q)}{q}H + \sum_{j=1}^{J} c(j,\chi_0)(\log H)^{1/2-j} + O\left((\log H)^{-1/2-J}\right). \tag{3.37}$$

Using the expansion of $D(s)\Gamma(s)$ above, we have

$$b(1/2) = a(3/2)\log q = -\frac{2K}{\sqrt{\pi}}\log q \implies c(1,\chi_0) = \frac{b(1/2)}{2K^2\Gamma(1/2)} = -\frac{1}{K\pi}\log q.$$

We now complete the proof of Theorem 3.1.2. Using (3.27) and (3.37) and the orthogonality relations, we have for $v \neq 0$,

$$\begin{aligned} S(q{,}v{,}H) &= \sum_{\substack{h\geq 1 \\ h\equiv v \,(\mathrm{mod}\, q)}} \mathfrak{S}(\{0{,}h\})e^{-h/H} = \frac{1}{\phi(q)}\sum_{\chi}\chi(v)^{-1}S(H,\chi) \\ &= \frac{1}{\phi(q)}S(H,\chi_0) + \frac{1}{2K^2\phi(q)}\sum_{\chi\neq\chi_0}\chi(v)^{-1}C_{q,\chi} + O(\exp(-c\sqrt{\log H})) \\ &= \frac{H}{q} + \frac{1}{2K^2\phi(q)}\sum_{\chi\neq\chi_0}\chi(v)^{-1}C_{q,\chi} + \sum_{j=1}^{J}\frac{c(j,\chi_0)}{\phi(q)}(\log H)^{1/2-j} + O\left((\log H)^{-1/2-J}\right). \end{aligned}$$

For $v = 0$, we use (3.32) and the above to get

$$\begin{aligned} S(q{,}0{,}H) &= S(H) - \sum_{v\in(\mathbb{Z}/q\mathbb{Z})^*} S(q{,}v{,}H) \\ &= \frac{H}{q} - \frac{2}{K\pi}\sqrt{\log\ H} + \sum_{j=1}^{J}\left(c(j) - c(j,\chi_0)\right)(\log H)^{1/2-j} + O\left((\log H)^{-1/2-J}\right), \end{aligned}$$

where we used the fact that

$$\sum_{v\in(\mathbb{Z}/q\mathbb{Z})^*}\sum_{\chi\neq\chi_0}\chi(v)^{-1}C_{q,\chi} = 0$$

by the orthogonality relations. This completes the proof of the proposition, with $c_1(j) = c(j,\chi_0)/\phi(q)$ and $c_0(j) = c(j) - c(j,\chi_0)$ for $j \geq 1$, from which the relation $c_0(j) + \phi(q)c_1(j) = c(j)$ easily follows. From the values $c(1)$ and $c(1,\chi_0)$ computed above, we have

$$c_1(1) = -\frac{\log q}{K\phi(q)\pi} \text{ and } c_0(1) = \frac{1}{K\pi}\left(\frac{\omega+\gamma}{2} + \log q\right), \tag{3.38}$$

where γ is the Euler-Mascheroni constant and ω is defined in Lemma 3.5.2. □

Lemma 3.5.2. *Let $Z(s)$ be the function defined by (3.34). Then,*

$$Z'(0) = \frac{K}{\sqrt{\pi}}\omega \approx -0.3851314513\ldots$$

where

$$\omega = \log \frac{2}{\pi^2} + \frac{L'(1,\chi_4)}{L(1,\chi_4)} + 2 \sum_{p\equiv 3(4)} \frac{\log p}{p^2-1}.$$

Proof. Firstly, observe $M(s)$ in (3.34) and rewrite it as

$$Z(s) = \zeta(s)M(s) = \zeta(s)A(s)B(s)$$

where

$$A(s) = 1 - 2^{-s} + 2^{-2s},$$
$$B(s) = \left(L(s+1,\chi_4)(1-2^{-(s+1)}) \prod_{p\equiv 3(4)} \left(1 - \frac{1}{p^{2(s+1)}}\right)\right)^{-1/2}.$$

Then, we have

$$\frac{Z'}{Z}(0) = \frac{\zeta'}{\zeta}(0) + \frac{A'}{A}(0) + \frac{B'}{B}(0),$$

Hence, we need to compute ζ, A, A', B and B' at $s = 0$. Indeed, the following special values for ζ are well-known:

$$\zeta(0) = -\frac{1}{2}, \quad \zeta'(0) = -\frac{\log 2\pi}{2}.$$

Moreover, for $A(s)$, we have $A(0) = 1, A'(0) = -\log 2$. We may use the recursive formula for $B(s)$ to obtain $B(0) = 4K/\sqrt{\pi}$ and

$$\frac{B'}{B}(0) = -\frac{1}{2}\left(\frac{L'(1,\chi_4)}{L(1,\chi_4)} + \log 2 + 2\sum_{p\equiv 3(4)} \frac{\log p}{p^2-1}\right) = -\frac{1}{2}\left(\log 2 + \alpha_1 + \beta_1\right)$$

where we denote

$$\alpha_1 = \frac{L'(1,\chi_4)}{L(1,\chi_4)}, = 0.2456096036\ldots, \quad \beta_1 = 2\sum_{p\equiv 3(4)} \frac{\log p}{p^2-1} = 0.4574727064\ldots.$$

One can compute the value of α_1 by $L(1,\chi_4) = \pi/4$ and $L'(1,\chi_4) = 0.192901331574902\ldots$. □

3.6 Proof of Theorem 3.3.4

In the heuristic leading to Conjecture 3.1.1, we used Theorem 3.3.4 to justify that the terms involving a sum of singular series for sets with three or more elements contribute to the error term. Theorem 3.3.4 is an analogue of [MS04, Theorem 2] of Montgomery and Soundararajan adapted from primes to sums of two squares. We now prove Theorem 3.3.4, following closely the argument developed in [MS04], without giving all the details but insisting on the points that are different in the case of the sums of two squares. To help with the comparison, we stay close to the notation used in loc. cit. so we may use notation that differs from the rest of the paper, which should not cause trouble to the reader as this section is relatively independent from the rest of the paper. We use the standard notation $e(x) = e^{2i\pi x}$.

3.6.1 The singular series

The first step in the proof is to write the singular series as an actual series (and not a Euler product), the way it was introduced by Hardy and Littlewood (see [MS04, Lemma 3]). We begin with giving a new expression for the local factors of the singular series. Let $\mathcal{D} = \{d_1, \dots, d_k\} \subseteq \mathbb{Z}$. We recall that for any $p \not\equiv 1 \pmod 4$, we have

$$\delta_{\mathcal{D}}(p) = \lim_{\alpha\to\infty} \frac{\#\{0 \leq a < p^\alpha : \forall d \in \mathcal{D}, a + d \equiv \square + \square \pmod{p^\alpha}\}}{p^\alpha},$$

and the singular series is defined by

$$\mathfrak{S}(\mathcal{D}) := \prod_{p\not\equiv 1 \pmod 4} \frac{\delta_{\mathcal{D}}(p)}{(\delta_{\{0\}}(p))^k}.$$

Lemma 3.6.1. *Let $\mathcal{D} = \{d_1, \dots, d_k\} \subseteq \mathbb{Z}$ be a set with k elements. For any prime number $p \not\equiv 1 \pmod 4$, one has*

$$\frac{\delta_{\mathcal{D}}(p)}{\delta_{\{0\}}(p)^k} = \sum_{q_1,\dots,q_k \mid p^\infty} \prod_{i=1}^k \frac{\lambda_2(q_i)}{q_i} A_{\mathcal{D}}(q_1, \dots, q_k),$$

where for any $q_1, \dots, q_k \in \mathbb{N}$,

$$A_{\mathcal{D}}(q_1, \dots, q_k) = \sum_{\substack{a_1,\dots,a_k \\ 1\leq a_i\leq q_i, (q_i,a_i)=1 \\ \sum_{i=1}^k \frac{a_i}{q_i}\in\mathbb{Z}}} e\left(\sum_{i=1}^k \frac{a_i d_i}{q_i}\right) \prod_{i=1}^k C(q_i,a_i),$$

with

$$C(q,a) = \begin{cases} 1 & \text{if } q \text{ is odd,} \\ 0 & \text{if } 2 \mid q \text{ but } 4 \nmid q, \\ 2e(-a/4) & \text{if } 4 \mid q. \end{cases}$$

and λ_2 is the multiplicative function defined on the prime powers by

$$\lambda_2(p^m) = \begin{cases} (-1)^m & \text{if } p \text{ is odd,} \\ 1 & \text{if } p = 2. \end{cases}$$

Proof. Let $p \equiv 3 \pmod 4$ be a prime number. For $\mathcal{D} = \{d_1, \dots, d_k\} \subseteq \mathbb{Z}$ a set with k elements, we deduce from [FKR17, Proposition 5.1, Proposition 5.3(a) and (5.4)] that

$$\delta_{\mathcal{D}}(p) = \lim_{\alpha\to\infty} p^{-\alpha} \sum_{x=1}^{p^\alpha} \prod_{i=1}^{k} \mathbf{1}_{S_{p,\alpha}}(x+d_i)$$

where $\mathbf{1}_{S_{p,\alpha}}$ is the characteristic function of the set $S_{p,\alpha} = \{p^{2\beta}m : 0 \leq \beta < \frac{\alpha}{2}, m \not\equiv 0 \,(\mathrm{mod}\, p)\}$. In particular, for α even, following the idea of the proof of [MV86, Lemma 2], we write that

$$\begin{aligned}
\mathbf{1}_{S_{p,\alpha}}(x) &= \sum_{\beta=0}^{\frac{\alpha}{2}-1} \sum_{s|p} \frac{\mu(s)}{p^{2\beta}s} \sum_{a=1}^{p^{2\beta}s} e\left(\frac{ax}{p^{2\beta}s}\right) \\
&= \sum_{\beta=0}^{\frac{\alpha}{2}-1} \frac{1}{p^{2\beta}} \left\{ \left(1-\frac{1}{p}\right) \sum_{r|p^{2\beta}} \sum_{a\in(\mathbb{Z}/r\mathbb{Z})^*} e\left(\frac{ax}{r}\right) - \frac{1}{p} \sum_{a\in(\mathbb{Z}/p^{2\beta+1}\mathbb{Z})^*} e\left(\frac{ax}{p^{2\beta+1}}\right) \right\} \\
&= \sum_{\gamma=0}^{\alpha-2} \left\{ \sum_{\beta=\lceil\frac{\gamma}{2}\rceil}^{\frac{\alpha}{2}-1} \frac{1}{p^{2\beta}} \left(1-\frac{1}{p}\right) \sum_{a\in(\mathbb{Z}/p^\gamma\mathbb{Z})^*} e\left(\frac{ax}{p^\gamma}\right) \right\} - \sum_{\beta=0}^{\frac{\alpha}{2}-1} \frac{1}{p^{2\beta+1}} \sum_{a\in(\mathbb{Z}/p^{2\beta+1}\mathbb{Z})^*} e\left(\frac{ax}{p^{2\beta+1}}\right) \\
&= \sum_{\gamma=0}^{\alpha-1} \frac{(-p)^{-\gamma} - p^{-\alpha}}{1+\frac{1}{p}} \sum_{a\in(\mathbb{Z}/p^\gamma\mathbb{Z})^*} e\left(\frac{ax}{p^\gamma}\right),
\end{aligned}$$

where we used the fact that α is even in the last line. Thus, we have,

$$\frac{\delta_{\mathcal{D}}(p)}{\delta_{\{0\}}(p)^k} = \lim_{\substack{\alpha\to\infty \\ \alpha \text{ even}}} p^{-\alpha} \sum_{x=1}^{p^\alpha} \prod_{i=1}^{k} \left(\sum_{\gamma_i=0}^{\alpha-1} ((-p)^{-\gamma_i} - p^{-\alpha}) \sum_{a_i\in(\mathbb{Z}/p^{\gamma_i}\mathbb{Z})^*} e\left(\frac{(x+d_i)a_i}{p^{\gamma_i}}\right) \right),$$

since $\delta_{\{0\}}(p) = (1+\frac{1}{p})^{-1}$ by [FKR17, Proposition 5.3(c)]. We swap the sums and begin with the sum over x, as $\gamma_1, \ldots, \gamma_k \leq \alpha - 1$, we have

$$p^{-\alpha} \sum_{x=1}^{p^\alpha} e\left(x \sum_{i=1}^{k} \frac{a_i}{p^{\gamma_i}}\right) = \begin{cases} 1 & \text{if } \sum_{i=1}^{k} \frac{a_i}{p^{\gamma_i}} \in \mathbb{Z} \\ 0 & \text{otherwise.} \end{cases}$$

This yields

$$\frac{\delta_{\mathcal{D}}(p)}{\delta_{\{0\}}(p)^k} = \lim_{\substack{\alpha\to\infty \\ \alpha \text{ even}}} \sum_{\gamma_1=0}^{\alpha-1} \cdots \sum_{\gamma_k=0}^{\alpha-1} \prod_{i=1}^{k} ((-p)^{-\gamma_i} - p^{-\alpha}) A_{\mathcal{D}}(p^{\gamma_1}, \ldots, p^{\gamma_k}).$$

We obtain the formula announced in the Lemma for $p \equiv 3 \,(\mathrm{mod}\, 4)$ by taking the limit $\alpha \to \infty$, and using the bound $|A_{\mathcal{D}}(q_1, \ldots, q_k)| \leq \frac{q_1 \cdots q_k}{[q_1, \ldots, q_k]}$ (see (3.46)).

The proof is similar for $p = 2$. By [FKR17, Proposition 5.2(a) and (5.3)], for $\alpha \geq 2$ we can take $S_{2,\alpha} = \{2^\beta m : 0 \leq \beta < \alpha - 1, m \equiv 1 \,(\mathrm{mod}\, 4)\}$, and [FKR17, Proposition 5.2(c)] gives $\delta_{\{0\}}(2) = \frac{1}{2}$.

We write that

$$\begin{aligned}
\mathbf{1}_{S_{2,\alpha}}(x) &= \sum_{\beta=0}^{\alpha-2} 2^{-\beta-2} \sum_{a=1}^{2^\beta} e\left(\frac{ax}{2^\beta}\right) \sum_{t=1}^{4} e\left((\tfrac{x}{2^\beta}-1)\frac{t}{4}\right) \\
&= \sum_{\beta=0}^{\alpha-2} 2^{-\beta-2} \sum_{r|2^{\beta+2}} \sum_{b\in(\mathbb{Z}/r\mathbb{Z})^*} e\left(\frac{(x-2^\beta)b}{r}\right) \\
&= \sum_{\gamma=0}^{\alpha} \sum_{b\in(\mathbb{Z}/2^\gamma\mathbb{Z})^*} e\left(\frac{xb}{2^\gamma}\right) \sum_{\beta=\max\{0,\gamma-2\}}^{\alpha-2} 2^{-\beta-2} e\left(-2^{\beta-\gamma}b\right) \\
&= \sum_{\gamma=0}^{\alpha} \sum_{b\in(\mathbb{Z}/2^\gamma\mathbb{Z})^*} e\left(\frac{xb}{2^\gamma}\right) \left(C(2^\gamma,b)2^{-\gamma-1} - 2^{-\alpha}\right),
\end{aligned}$$

note that in the sum we always have $(b,2)=1$ so $1+e(\frac{-b}{2})=0$. Thus, we have

$$\begin{aligned}
\frac{\delta_{\mathcal{D}}(2)}{\delta_{\{0\}}(2)^k} &= \lim_{\alpha\to\infty} 2^{-\alpha+k} \sum_{x=1}^{2^\alpha} \prod_{i=1}^{k} \mathbf{1}_{S_{2,\alpha}}(x+d_i) \\
&= \lim_{\alpha\to\infty} 2^{-\alpha} \sum_{x=1}^{2^\alpha} \prod_{i=1}^{k} \left(\sum_{\gamma_i=0}^{\alpha} \sum_{b_i\in(\mathbb{Z}/2^{\gamma_i}\mathbb{Z})^*} e\left(\frac{(x+d_i)b_i}{2^{\gamma_i}}\right) \left(C(2^{\gamma_i},b_i)2^{-\gamma_i} - 2^{-\alpha+1}\right) \right)
\end{aligned}$$

Exchanging the sums and computing the sum over x first, this yields

$$\frac{\delta_{\mathcal{D}}(2)}{\delta_{\{0\}}(2)^k} = \lim_{\alpha\to\infty} \sum_{\gamma_1=0}^{\alpha} \cdots \sum_{\gamma_k=0}^{\alpha} \prod_{i=1}^{k} 2^{-\gamma_i} A_{\mathcal{D}}(2^{\gamma_1},\ldots,2^{\gamma_k}),$$

which gives the formula announced in the Lemma for $p=2$. □

We now give the analogue of [MS04, (44) and (45)]. The main difference between the case of primes and the case of sum of two squares is that the local probabilities $\delta_{\mathcal{D}}(p)$ at each prime p involve all powers of p, and then the sum over $q_1,\ldots,q_k$ in Lemma 3.6.1 runs over all integers (and not only square-free integers). We then approximate $\mathfrak{S}(\mathcal{D})$ by taking all integers supported on primes $p\le y$ and appearing with power at most N, for the appropriate values of y and N.

Lemma 3.6.2. *Let $\mathcal{D}\subseteq\mathbb{N}\cap[1,h]$ be a set with k elements. Let $y>h$, $N\ge 4\log y$, and $P_y := \prod_{\substack{p\le y\\ p\not\equiv 1 \,(\mathrm{mod}\,4)}} p$. Then,*

$$\mathfrak{S}(\mathcal{D}) = \sum_{q_1,\ldots,q_k|P_y^N} \prod_{i=1}^{k} \frac{\lambda_2(q_i)}{q_i} A_{\mathcal{D}}(q_1,\ldots,q_k) + O_k(y^{-1}(\log y)^{k-1}) \tag{3.39}$$

and

$$\mathfrak{S}_0(\mathcal{D}) = \sum_{\substack{q_1,\ldots,q_k|P_y^N\\ q_i>1}} \prod_{i=1}^{k} \frac{\lambda_2(q_i)}{q_i} A_{\mathcal{D}}(q_1,\ldots,q_k) + O_k(y^{-1}(\log y)^{k-1}), \tag{3.40}$$

where $A_{\mathcal{D}}(q_1,\dots,q_k)$ is defined in Lemma 3.6.1.

Proof. First, it follows from the Chinese Remainder Theorem that for $q_1,\dots,q_k,q_1',\dots,q_k'\in\mathbb{N}$ satisfying $\left(\prod_{i=1}^k q_i,\prod_{i=1}^k q_i'\right)=1$, one has $A_{\mathcal{D}}(q_1,\dots,q_k)A_{\mathcal{D}}(q_1',\dots,q_k')=A_{\mathcal{D}}(q_1q_1',\dots,q_1q_1')$. Since $y>h\geq\max\mathcal{D}$, from [FKR17, Proposition 5.3.(c)] we deduce

$$\prod_{p>y}\frac{\delta_{\mathcal{D}}(p)}{\delta_{\{0\}}(p)^k}=\prod_{p>y}(1+\tfrac{1}{p})^{k-1}(1-\tfrac{k-1}{p})=\prod_{p>y}(1+O_k(p^{-2}))=1+O_k((y\log y)^{-1}).$$

By definition we have $\delta_{\mathcal{D}}(p)\leq 1$ for all prime number p, thus

$$\prod_{p\leq y}\frac{\delta_{\mathcal{D}}(p)}{\delta_{\{0\}}(p)^k}\leq 2^k\prod_{\substack{p\leq y\\ p\equiv 3\,(\mathrm{mod}\,4)}}(1+\tfrac{1}{p})^k\ll_k(\log y)^k,$$

which gives

$$\mathfrak{S}(\mathcal{D})=\prod_{p\leq y}\frac{\delta_{\mathcal{D}}(p)}{\delta_{\{0\}}(p)^k}+O(y^{-1}(\log y)^{k-1}).$$

Using Lemma 3.6.1 and the bound $|A_{\mathcal{D}}(q_1,\dots,q_k)|\leq 2^{\frac{k}{2}}\frac{q_1\dots q_k}{[q_1,\dots,q_k]}$ (see (3.46)), we have

$$\begin{aligned}\frac{\delta_{\mathcal{D}}(p)}{\delta_{\{0\}}(p)^k}&=\sum_{q_1,\dots,q_k|p^N}\prod_{i=1}^k\frac{\lambda_2(q_i)}{q_i}A_{\mathcal{D}}(q_1,\dots,q_k)+O_k\Big(\sum_{p^{N+1}|q_1|p^\infty}\sum_{q_2,\dots,q_k|p^\infty}\frac{1}{[q_1,\dots,q_k]}\Big)\\&=\sum_{q_1,\dots,q_k|p^N}\prod_{i=1}^k\frac{\lambda_2(q_i)}{q_i}A_{\mathcal{D}}(q_1,\dots,q_k)+O_k\Big(\sum_{n=N+1}^{\infty}\frac{(n+1)^{k-1}}{p^n}\Big)\\&=\sum_{q_1,\dots,q_k|p^N}\prod_{i=1}^k\frac{\lambda_2(q_i)}{q_i}A_{\mathcal{D}}(q_1,\dots,q_k)+O_k(p^{-N-1}(N+2)^{k-1}).\end{aligned}$$

Moreover, using again the bound (3.46), we have

$$\begin{aligned}\prod_{p\leq y}\sum_{q_1,\dots,q_k|p^N}\prod_{i=1}^k\frac{\lambda_2(q_i)}{q_i}A_{\mathcal{D}}(q_1,\dots,q_k)&\ll_k\sum_{q_1,\dots,q_k|P_y^N}\prod_{i=1}^k\frac{|\lambda_2(q_i)|}{q_i}\frac{q_1\cdot\dots\cdot q_k}{[q_1,\dots,q_k]}\\&\leq\sum_{q_1,\dots,q_k|P_y^N}\frac{1}{[q_1,\dots,q_k]}\\&\ll_k\prod_{p\leq y}\sum_{n=0}^{N}\frac{(n+1)^k}{p^n}\leq\prod_{p\leq y}(1+\frac{C_k}{p})\ll_\varepsilon y^\varepsilon\end{aligned}\qquad(3.41)$$

for some constant $C_k > 0$, for any $\varepsilon > 0$. Finally,

$$\prod_{p\leq y}\frac{\delta_{\mathcal{D}}(p)}{\delta_{\{0\}}(p)^k} = \sum_{q_1,\dots,q_k|P_y^N}\prod_{i=1}^{k}\frac{\lambda_2(q_i)}{q_i}A_{\mathcal{D}}(q_1,\dots,q_k) + O_{k,\varepsilon}\Big(y^{\varepsilon}\sum_{\substack{q|P_y\\ q\neq 1}} q^{-N-1}N^{(k-1)\omega(q)}\Big)$$

$$= \sum_{q_1,\dots,q_k|P_y^N}\prod_{i=1}^{k}\frac{\lambda_2(q_i)}{q_i}A_{\mathcal{D}}(q_1,\dots,q_k) + O_{k,\varepsilon}(y^{\varepsilon}2^{-N}N^{k-1}).$$

Choosing $\frac{\log y}{\log 2}(1+\varepsilon) < N$ gives (3.39). We deduce (3.40) from (3.39) using the formula

$$\mathfrak{S}_0(\mathcal{D}) = \sum_{\mathcal{T}\subseteq\mathcal{D}}(-1)^{|\mathcal{D}\setminus\mathcal{T}|}\mathfrak{S}(\mathcal{T})$$

and the relation $A_{\{d_1,\dots,d_k\}}(q_1,\dots,q_{k-1},1) = A_{\{d_1,\dots,d_{k-1}\}}(q_1,\dots,q_{k-1})$. □

In particular, taking $y = h^{k+1}$ in (3.40), one has

$$\sum_{\substack{\mathcal{D}\subseteq[1,h]\\ |\mathcal{D}|=k}}\mathfrak{S}_0(\mathcal{D}) = \sum_{\substack{q_1,\dots,q_k|P_y^N\\ q_i>1}}\prod_{i=1}^{k}\frac{\lambda_2(q_i)}{q_i}\sum_{\substack{\mathcal{D}\subseteq[1,h]\\ |\mathcal{D}|=k}}A_{\mathcal{D}}(q_1,\dots,q_k) + o_k(1). \tag{3.42}$$

3.6.2 An easier version of the main term

To continue with notation similar to [MS04], we define

$$V_k(y,N,h) = \sum_{\substack{q_1,\dots,q_k|P_y^N\\ q_i>1}}\prod_{i=1}^{k}\frac{\lambda_2(q_i)}{q_i}\sum_{1\leq d_1,\dots,d_k\leq h}A_{(d_1,\dots,d_k)}(q_1,\dots,q_k), \tag{3.43}$$

where we remark that the difference with the main term above is that $d_1,\dots,d_k$ do not have to be distinct. Let us introduce some other useful notations and results from [MS04]. For $\alpha\in\mathbb{R}$, we denote

$$E_h(\alpha) = \sum_{d=1}^{h}e(\alpha d) \quad\text{and}\quad F_h(\alpha) = \min(h,\|\alpha\|^{-1}), \tag{3.44}$$

where $\|\cdot\|$ is the distance to the nearest integer, so that we have $|E_h(\alpha)| \leq F_h(\alpha)$. We have (see [MS04, (54)])

$$\sum_{a=1}^{q-1}F_h(\tfrac{a}{q})^2 \ll q\min(q,h). \tag{3.45}$$

We will also use the following result from the work of Montgomery and Vaughan [MV89] and which is an analogue of [MS04, Lemma 1] that applies to the case of non necessarily square-free numbers.

Lemma 3.6.3 (Theorem 1 of [MV89]). *Let $k \geq 2$ be an integer and for $1 \leq i \leq k$, let $q_i \in \mathbb{N}$ and G_i be a 1-periodic complex valued function. Then, we have*

$$\Big| \sum_{\substack{a_1,\dots,a_k \\ 1\leq a_i\leq q_i,(q_i,a_i)=1 \\ \sum_{i=1}^k \frac{a_i}{q_i}\in\mathbb{Z}}} \prod_{i=1}^k G_i(\tfrac{a_i}{q_i})\Big| \leq \frac{1}{[q_1,\dots,q_k]} \prod_{i=1}^k \big(q_i \sum_{\substack{1\leq a_i\leq q_i \\ (q_i,a_i)=1}} |G_i(\tfrac{a_i}{q_i})|^2\big)^{\frac{1}{2}}.$$

In particular we deduce the bound for $A_{\mathcal{D}}(q_1,\dots,q_k)$ that we used in the proofs of Lemma 3.6.1 and 3.6.2:

$$|A_{\mathcal{D}}(q_1,\dots,q_k)| \leq \frac{1}{[q_1,\dots,q_k]} \prod_{i=1}^k \big(q_i \sum_{\substack{1\leq a_i\leq q_i \\ (q_i,a_i)=1}} |C(q_i,a_i)|^2\big)^{\frac{1}{2}}. \tag{3.46}$$

We also have a bound for $V_k(y,N,h)$.

Corollary 1. For any $h,y,N>0$, one has $V_k(y,N,h) \ll_{k,\varepsilon} h^{\frac{k}{2}} y^{\varepsilon}$.

Proof. Recall that we defined

$$V_k(y,N,h) = \sum_{\substack{q_1,\dots,q_k | P_y^N \\ q_i>1}} \prod_{i=1}^k \frac{\lambda_2(q_i)}{q_i} \sum_{\substack{a_1,\dots,a_k \\ 1\leq a_i\leq q_i,(q_i,a_i)=1 \\ \sum_{i=1}^k \frac{a_i}{q_i}\in\mathbb{Z}}} \prod_{i=1}^k E_h\Big(\frac{a_i d_i}{q_i}\Big) C(q_i,a_i),$$

where $C(q,a)=1$ for odd q and $|C(q,a)| \leq 2$ in general. We use (3.44) to write

$$|V_k(y,N,h)| \leq \sum_{\substack{q_1,\dots,q_k | P_y^N \\ q_i>1}} \prod_{i=1}^k \frac{2}{q_i} \sum_{\substack{a_1,\dots,a_k \\ 1\leq a_i\leq q_i,(q_i,a_i)=1 \\ \sum_{i=1}^k \frac{a_i}{q_i}\in\mathbb{Z}}} \prod_{i=1}^k F_h(\tfrac{a_i}{q_i}).$$

Then Lemma 3.6.3, (3.45) and the bound in (3.41) yield

$$\begin{aligned} |V_k(y,N,h)| &\leq \sum_{\substack{q_1,\dots,q_k | P_y^N \\ q_i>1}} \frac{2^k}{[q_1,\dots,q_k]} \prod_{i=1}^k \frac{1}{q_i^{\frac{1}{2}}} \Big(\sum_{\substack{1\leq a_i\leq q_i \\ (q_i,a_i)=1}} F_h(\tfrac{a_i}{q_i})^2 \Big)^{\frac{1}{2}} \\ &\leq \sum_{\substack{q_1,\dots,q_k | P_y^N \\ q_i>1}} \frac{2^k}{[q_1,\dots,q_k]} h^{\frac{k}{2}} \ll_{k,\varepsilon} h^{\frac{k}{2}} y^{\varepsilon}, \end{aligned}$$

which is the bound announced. ∎

3.6.3 The main estimate

We now prove the analogue of [MS04, (60)], writing $\sum_{\substack{\mathcal{D}\subseteq[1,h]\\|\mathcal{D}|=k}} \mathfrak{S}_0(\mathcal{D})$ in terms of $V_k(y,N,h)$. Again, the idea of the proof is very similar to the work of Montgomery and Soundararajan, except that we deal with a wider summation (namely a sum over all integers instead of a sum over square-free integers).

Lemma 3.6.4. *For any $h > k \in \mathbb{N}$, let $y = h^{k+1}$ and $N \geq 4\log y$. One has*

$$\sum_{\substack{\mathcal{D}\subseteq[1,h]\\|\mathcal{D}|=k}} \mathfrak{S}_0(\mathcal{D}) = \sum_{j=0}^{k/2} \binom{k}{2j} \frac{(2j)!}{j!2^j} \left(-h \sum_{1<d|P_y^N} \frac{C(d)\phi(d)}{d^2} \right)^j V_{k-2j}(y,N,h) + O_{k,\varepsilon}(h^{\frac{k-1}{2}} y^{\varepsilon}),$$

where $V_k(y,N,h)$ is defined in (3.43), *and*

$$C(d) = \begin{cases} 1 \text{ if } d \text{ is odd} \\ 0 \text{ if } 2 \mid q, 4 \nmid d \\ 4 \text{ if } 4 \mid d. \end{cases}$$

Proof. Following the arguments of [MS04], we can prove the analogue of [MS04, (52)] in our context, which is

$$\sum_{\substack{\mathcal{D}\subseteq[1,h]\\|\mathcal{D}|=k}} A_{\mathcal{D}}(q_1,\dots,q_k) = \sum_{\mathcal{P}=\{\mathcal{S}_1,\dots,\mathcal{S}_M\}} w(\mathcal{P}) \sum_{\substack{a_1,\dots,a_k\\1\leq a_i\leq q_i,(q_i,a_i)=1\\ \sum_{i=1}^k \frac{a_i}{q_i}\in\mathbb{Z}}} \prod_{i=1}^k C(q_i,a_i) \prod_{m=1}^M \sum_{d_m=1}^h e\left(\sum_{i\in\mathcal{S}_m} \frac{a_i}{q_i} d_m \right), \tag{3.47}$$

where the first sum is over partitions $\mathcal{P} = \{\mathcal{S}_1,\dots,\mathcal{S}_M\}$ of $\{1,\dots,k\}$, and $w(\mathcal{P})$ is defined in [MS04, p. 17].

In the case of a partition $\mathcal{P}$ containing at least one part of size ≥ 3, write $\mathcal{N}_1 = \bigcup_{|\mathcal{S}_m|=1} \mathcal{S}_m$, $\mathcal{N}_2 = \{1,\dots,k\} \setminus \mathcal{N}_1$ and $m_2 = |\{1 \leq m \leq M : |\mathcal{S}_m| \geq 2\}|$. Using (3.44) and $|C(q,x)| \leq 2$, we have

$$\Bigg| \sum_{\substack{a_1,\dots,a_k\\1\leq a_i\leq q_i,(q_i,a_i)=1\\ \sum_{i=1}^k \frac{a_i}{q_i}\in\mathbb{Z}}} \prod_{i=1}^k C(q_i,a_i) \prod_{m=1}^M \sum_{d_m=1}^h e\left(\sum_{i\in\mathcal{S}_m} \frac{a_i}{q_i} d_m \right) \Bigg| \leq 2^k h^{m_2} \sum_{\substack{a_1,\dots,a_k\\1\leq a_i\leq q_i,(q_i,a_i)=1\\ \sum_{i=1}^k \frac{a_i}{q_i}\in\mathbb{Z}}} \prod_{i\in\mathcal{N}_1} F_h(\tfrac{a_i}{q_i}).$$

Then we apply Lemma 3.6.3 and the bound (3.45) to obtain that the sum above is

$$\begin{aligned} &\leq \frac{2^k h^{m_2}}{[q_1,\dots,q_k]} \prod_{i\in\mathcal{N}_1} \Big(q_i \sum_{\substack{1\leq a_i\leq q_i\\(q_i,a_i)=1}} (F_h(\tfrac{a_i}{q_i}))^2 \Big)^{\frac12} \prod_{i\in\mathcal{N}_2} \Big(q_i \sum_{\substack{1\leq a_i\leq q_i\\(q_i,a_i)=1}} 1^2 \Big)^{\frac12} \\ &\leq 2^k h^{\frac{k-1}{2}} \frac{q_1\cdot\dots\cdot q_k}{[q_1,\dots,q_k]}, \end{aligned}$$

where we used $\frac{1}{2}|\mathcal{N}_1| + m_2 \le \frac{k-1}{2}$ when the partition $\mathcal{P}$ contains at least one part of size ≥ 3. Replacing this bound in (3.47) and then in (3.42), we sum over $1 < q_1, \dots, q_k \mid P_y^N$ as in (3.42) and use the bound (3.41) to obtain that the contribution of the partitions containing at least one part of size ≥ 3 in $\sum_{\substack{\mathcal{D}\subseteq[1,h]\\|\mathcal{D}|=k}} \mathfrak{S}_0(\mathcal{D})$ is at most $O_{k,\varepsilon}(h^{\frac{k-1}{2}+\varepsilon})$.

We now turn our attention to partitions of $\{1,\dots,k\}$ with sets of size at most 2. The combinatorics leading to [MS04, (56)] work similarly and give

$$\begin{aligned}\sum_{\substack{\mathcal{D}\subseteq[1,h]\\|\mathcal{D}|=k}} \mathfrak{S}_0(\mathcal{D}) = \sum_{0\le j\le\frac{k}{2}} (-1)^j \binom{k}{2j} \frac{(2j)!}{j!2^j} \sum_{r_1,\dots,r_j|P_y^N} \sum_{\substack{b_1,\dots,b_j\\1\le b_i\le r_i,(r_i,b_i)=1}} \prod_{i=1}^{j} H(\tfrac{b_i}{r_i})\\ \times \sum_{\substack{q_{2j+1},\dots,q_k|P_y^N\\q_i>1}} \sum_{\substack{a_{2j+1},\dots,a_k\\1\le a_i\le q_i,(q_i,a_i)=1\\\sum_{i=1}^{j}\frac{b_i}{r_i}+\sum_{i=2j+1}^{k}\frac{a_i}{q_i}\in\mathbb{Z}}} \prod_{i=2j+1}^{k} \frac{\lambda_2(q_i)C(q_i,a_i)}{q_i} \sum_{d_i=1}^{h} e\left(\frac{a_i}{q_i}d_i\right)\\ + O_k(h^{\frac{k-1}{2}+\varepsilon})\end{aligned} \tag{3.48}$$

where

$$H(\tfrac{b}{r}) = \sum_{\substack{q_1,q_2|P_y^N\\q_i>1}} \sum_{\substack{a_1,a_2\\1\le a_i\le q_i,(q_i,a_i)=1\\\frac{a_1}{q_1}+\frac{a_2}{q_2}\in\frac{b}{r}+\mathbb{Z}}} \frac{\lambda_2(q_1)C(q_1,a_1)\lambda_2(q_2)C(q_2,a_2)}{q_1q_2} \sum_{d=1}^{h} e\left(\frac{b}{r}d\right).$$

In particular, we have

$$H(1) = \sum_{1<q|P_y^N} \sum_{\substack{1\le a\le q\\(q,a)=1}} \frac{C(q,a)C(q,q-a)}{q^2} h = h \sum_{1<q|P_y^N} \frac{C(q)\phi(q)}{q^2},$$

and the contribution of the terms with all $r_i = 1$ in the sum above is

$$\sum_{0\le j\le\frac{k}{2}} \binom{k}{2j} \frac{(2j)!}{j!2^j} \Big(-h \sum_{1<q|P_y^N} \frac{C(q)\phi(q)}{q^2} \Big)^j V_{k-2j}(y,N,h).$$

We now show that the contribution to (3.48) of the terms where not all r_i are 1 can be absorbed in the error term. Let ℓ be the number of i's for which $r_i > 1$. For any $\ell > 0$, and any $r_1, \ldots, r_\ell, q_{2j+1}, \ldots, q_k > 1$ up to re-ordering and applying Lemma 3.6.3, we have

$$\begin{aligned}\sum_{\substack{b_1,\ldots,b_\ell \\ 1\le b_i\le r_i,(r_i,b_i)=1}} \prod_{i=1}^{\ell} H\big(\frac{b_i}{r_i}\big) \sum_{\substack{a_{2j+1},\ldots,a_k \\ 1\le a_i\le q_i,(q_i,a_i)=1 \\ \sum_{i=1}^{j}\frac{b_i}{r_i}+\sum_{i=2j+1}^{k}\frac{a_i}{q_i}\in\mathbb{Z}}} \prod_{i=2j+1}^{k} \frac{\lambda_2(q_i)C(q_i,a_i)}{q_i} \sum_{d_i=1}^{h} e\left(\frac{a_i}{q_i}d_i\right) \\ \ll_k \frac{1}{[r_1,\ldots,r_\ell,q_{2j+1},\ldots,q_k]} \prod_{i=1}^{\ell}\Big(r_i \sum_{1\le b\le r_i,(r_i,b)=1} \Big|H\big(\frac{b}{r_i}\big)\Big|^2\Big)^{\frac{1}{2}} \\ \times \prod_{i=2j+1}^{k}\Big(\frac{1}{q_i}\sum_{1\le a\le q_i,(q_i,a)=1}\Big|F_h\big(\frac{a}{q_i}\big)\Big|^2\Big)^{\frac{1}{2}}.\end{aligned} \tag{3.49}$$

To obtain a bound for $\left|H\big(\frac{b}{r}\big)\right|$ we proceed similarly to [MS04] which gives

$$\begin{aligned} H\big(\tfrac{b}{r}\big) &\ll F_h\big(\tfrac{b}{r}\big) \sum_{\substack{s_1,s_2|P_y^N \\ [s_1,s_2]=r}} \sum_{\substack{c_1,c_2 \\ 1\le c_i\le s_i,(s_i,c_i)=1 \\ \frac{c_1}{s_1}+\frac{c_2}{s_2}\in\frac{b}{r}+\mathbb{Z}}} \frac{1}{s_1s_2} \sum_{\substack{t|P_y^N \\ (t,r)=1}} \frac{\phi(t)}{t^2} \\ &\ll F_h\big(\tfrac{b}{r}\big)\frac{1}{\phi(r)} \sum_{\substack{s_1,s_2|P_y^N \\ [s_1,s_2]=r}} \frac{\phi(s_1)\phi(s_2)}{s_1s_2} \prod_{\substack{p\le y \\ p\nmid r}}\Big(1+\frac{1}{1-p^{-1}}\sum_{n=1}^{N}p^{-n}\Big) \\ &\ll F_h\big(\tfrac{b}{r}\big)\frac{1}{r}\prod_{p|r}(1-p^{-1})^2(1+v_p(r)-v_p(r)p^{-1}) \prod_{\substack{p\le y \\ p\nmid r}}\Big(1+p^{-1}\Big) \\ &\ll F_h\big(\tfrac{b}{r}\big)\frac{d(r)}{r}\log y, \end{aligned}$$

where $d(r)$ is the number of divisors of r. Using (3.45), this gives

$$\sum_{\substack{1\le b\le r \\ (r,b)=1}} \left|H\big(\tfrac{b}{r}\big)\right|^2 \le \min(r,h)\frac{d(r)^2}{r}(\log y)^2.$$

Using this bound and (3.45) in (3.49) summed over all $r_1, \ldots, r_\ell, q_{2j+1}, \ldots, q_k > 1$ divisors of

P_y^N, we obtain

$$\sum_{\substack{r_1,\dots,r_\ell|P_y^N\\ r_i>1}} \sum_{\substack{b_1,\dots,b_\ell\\ 1\le b_i\le r_i,(r_i,b_i)=1}} \prod_{i=1}^{\ell} H(\tfrac{b_i}{r_i}) \sum_{\substack{q_{2j+1},\dots,q_k|P_y^N\\ q_i>1}} \sum_{\substack{a_{2j+1},\dots,a_k\\ 1\le a_i\le q_i,(q_i,a_i)=1\\ \sum_{i=1}^{j}\frac{y_i}{r_i}+\sum_{i=2j+1}^{k}\frac{a_i}{q_i}\in\mathbb{Z}}} \prod_{i=2j+1}^{k} \frac{\lambda_2(q_i)C(q_i,a_i)}{q_i} \sum_{d_i=1}^{h} e\left(\frac{a_i}{q_i}d_i\right)$$

$$\ll_k \sum_{\substack{r_1,\dots,r_\ell|P_y^N\\ r_i>1}} \sum_{\substack{q_{2j+1},\dots,q_k|P_y^N\\ q_i>1}} \frac{1}{[r_1,\dots,r_\ell,q_{2j+1},\dots,q_k]} \prod_{i=1}^{\ell}\Big(h^{\frac12}d(r_i)\log y\Big)h^{\frac{k-2j}{2}}$$

$$\ll_k h^{\frac{k+\ell-2j}{2}}(\log y)^\ell \sum_{m|P_y^N}\frac{1}{m}\Big(\sum_{r|m}d(r)\Big)^\ell\Big(\sum_{q|m}1\Big)^{k-2j}$$

$$\ll_k h^{\frac{k+\ell-2j}{2}}(\log y)^\ell \prod_{p|P_y}\Big(1+\sum_{n=1}^{N}(n+1)^{k+\ell-2j}(\frac{n+2}{2})^\ell p^{-n}\Big) \ll_{k,\ell,j,\varepsilon} h^{\frac{k+\ell-2j}{2}}y^\varepsilon.$$

Finally, summing the contribution for each $\ell \ge 0$ yields

$$\sum_{\substack{\mathcal{D}\subseteq[1,h]\\ |\mathcal{D}|=k}} \mathfrak{S}_0(\mathcal{D}) = \sum_{j=0}^{k/2}\binom{k}{2j}\frac{(2j)!}{j!2^j}\left(-h\sum_{1<d|P_y^N}\frac{C(d)\phi(d)}{d^2}\right)^j V_{k-2j}(y,N,h)$$
$$+\sum_{j=0}^{k/2}\binom{k}{2j}\frac{(2j)!}{j!2^j}\sum_{\ell=1}^{j}\binom{j}{\ell}\left(h\sum_{1<d|P_y^N}\frac{C(d)\phi(d)}{d^2}\right)^{j-\ell} O(h^{\frac{k+\ell-2j}{2}}y^\varepsilon)+O_k(h^{(k-1+\varepsilon)/2}),$$

and using $\sum_{1<d|P_y^N}\frac{C(d)\phi(d)}{d^2} \ll_\varepsilon y^\varepsilon$, we deduce

$$\sum_{\substack{\mathcal{D}\subseteq[1,h]\\ |\mathcal{D}|=k}} \mathfrak{S}_0(\mathcal{D}) = \sum_{j=0}^{k/2}\binom{k}{2j}\frac{(2j)!}{j!2^j}\left(-h\sum_{1<d|P_y^N}\frac{C(d)\phi(d)}{d^2}\right)^j V_{k-2j}(y,N,h)+O_{k,\varepsilon}(h^{\frac{k-1}{2}}y^\varepsilon).$$

which completes the proof of Lemma 3.6.4. □

The proof of Theorem 3.3.4 is now relatively straightforward. Lemma 3.6.4 gives for any $h > k \in \mathbb{N}$, and $N \ge 4(k+1)\log h$ that

$$\sum_{\substack{\mathcal{D}\subseteq[1,h]\\ |\mathcal{D}|=k}} \mathfrak{S}_0(\mathcal{D}) = \sum_{j=0}^{k/2}\binom{k}{2j}\frac{(2j)!}{j!2^j}\left(-h\sum_{1<d|P_y^N}\frac{C(d)\phi(d)}{d^2}\right)^j V_{k-2j}(h^{k+1},N,h)+O_{k,\varepsilon}(h^{\frac{k-1}{2}+\varepsilon}).$$

Then the bound from Corollary 1 yields

$$\sum_{\substack{\mathcal{D}\subseteq[1,h]\\|\mathcal{D}|=k}}\mathfrak{S}_0(\mathcal{D}) \ll_{k,\varepsilon} \sum_{j=0}^{k/2}\binom{k}{2j}\frac{(2j)!}{j!2^j}\left(h\sum_{1<d|P_y^N}\frac{C(d)\phi(d)}{d^2}\right)^j h^{\frac{k-2j}{2}+\varepsilon}+O_{k,\varepsilon}(h^{\frac{k-1}{2}+\varepsilon})$$
$$\ll_k h^{\frac{k}{2}+\varepsilon}\Big(\prod_{p<y}1-\frac{1}{p}\Big)^{\frac{-k}{2}} \ll_{k,\varepsilon} h^{\frac{k}{2}+2\varepsilon},$$

which finishes the proof of Theorem 3.3.4.

3.7 Integral form and improved error terms

Using the methods of Section 3.5 and Theorem 3.3.4, we can obtain a more precise form of the averages of the Hardy–Littlewood constants for sums of two squares of [Smi13, Theorem 1.1] and [FKR17, Proposition 1.3] (in a special case) by exhibiting a secondary term. In order to see the secondary term, we need to express the results of Section 3.5 differently, as a closed-form expression which contains implicitly all the descending powers of $\log h$. We first prove that we can write such an asymptotic for the number of sums of two squares, with a square-root cancellation error term under the Riemann Hypothesis (Theorem 3.2.1). The argument for the proof of Theorem 3.2.1 is essentially due to Selberg and known to experts, it appeared as a mathoverflow post [(ht13], and an exercise in the book of Koukoulopoulos [Kou20, Exercise 13.7]. Note also the observation of Tenenbaum [Ten15, page 291] as well as the independent analogue result of Gorodetsky and Rodgers [GR21, Theorem B.1] inspired by [RB02]. With the same techniques, we then prove Proposition 3.7.1, which exhibits the secondary term for the average of the Hardy-Littlewood constants for 2-tuples of sums of two squares. The general case is Proposition 3.1.3 and it follows by using Theorem 3.3.4 to show that the average over k-tuples reduces to the average over 2-tuples.

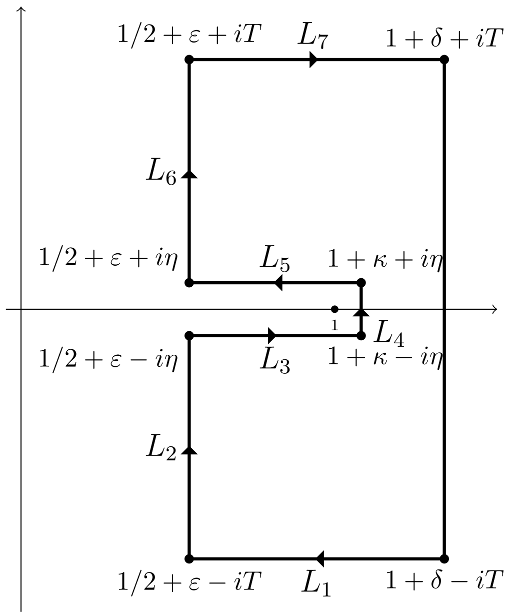
3.7.1 Proof of Theorem 3.2.1

We first assume the Riemann Hypothesis. Using Perron's formula, we have for any $\delta > 0$

$$\sum_{n\le x}1_{\mathbf{E}}(n) = \int_{1+\delta-iT}^{1+\delta+iT}F(s)\frac{x^s}{s}ds + O\Big(\frac{x^{1+\delta}\log x}{T}\Big), \tag{3.50}$$

where $F(s)=\zeta^{1/2}(s)L(s,\chi_4)^{1/2}(1-2^{-s})^{-1/2}\prod_{p\equiv 3\,(\mathrm{mod}\,4)}\left(1-p^{-2s}\right)^{-1/2}$ as seen in the proof of Theorem 3.2.4. The above path integral is part of a contour which encloses a region of analyticity of the integrand, which is the usual contour going from $1+\delta-iT$ to $1+\delta+iT$ then to $1/2+\varepsilon+iT$ then to $1/2+\varepsilon-iT$ and then back to $1+\delta-iT$ with a slit along the real axis between $1/2+\varepsilon$ and 1, with a line just above the real axis from $1/2+\varepsilon$ to 1, and a line just below the real axis from 1 to $1/2+\varepsilon$. More precisely, for any $\varepsilon,\eta>0$ and for $0<\kappa<\delta$, we define line segments L_j, $j=1,2,...,7$ as in Figure 3.2.

Together with the line segment $1+\delta+iT\to 1+\delta-iT$ of the integral (3.50), this gives the closed contour of Figure 3.2, which encloses a region of analyticity of the function $F(s)=G(s)(s-$



$$L_1 : 1+\delta-iT \to 1/2+\varepsilon-iT$$
$$L_2 : 1/2+\varepsilon-iT \to 1/2+\varepsilon-i\eta$$
$$L_3 : 1/2+\varepsilon-i\eta \to 1+\kappa-i\eta$$
$$L_4 : 1+\kappa-i\eta \to 1+\kappa+i\eta$$
$$L_5 : 1+\kappa+i\eta \to 1/2+\varepsilon+i\eta$$
$$L_6 : 1/2+\varepsilon+i\eta \to 1/2+\varepsilon+iT$$
$$L_7 : 1/2+\varepsilon+iT \to 1+\delta+iT.$$

Figure 3.2: The contour used in the proof of Theorem 3.2.1.

$1)^{-1/2}$ since we are assuming the Riemann Hypothesis and $\zeta^{1/2}(s)(s-1)^{1/2}$, $L(s,\chi_4)^{1/2}(1-2^{-s})^{-1/2}\prod_{p\equiv 3\,(\mathrm{mod}\,4)}(1-p^{-2s})^{-1/2}$ are analytic for $\mathrm{Re}(s) > 1/2+\varepsilon$. Then, using Cauchy's theorem, we have

$$\int_{1+\delta-iT}^{1+\delta+iT} F(s)\frac{x^s}{s}ds = \sum_{j=1}^{7}\int_{L_j} F(s)\frac{x^s}{s}ds.$$

The contribution coming from L_1,L_2,L_4,L_6,L_7 are bounded by the classical estimates, where we use the Lindelöf Hypothesis to bound $|\zeta^{1/2}(\sigma+it)|, |L^{1/2}(\sigma+it,\chi_4)| \ll_\sigma |t|^{\varepsilon_1}$ for $1/2 < \sigma < 1$ and $\varepsilon_1 > 0$. For the horizontal integral over L_1, we have

$$\int_{L_1} F(s)\frac{x^s}{s}ds \ll \int_{1/2+\varepsilon}^{1+\delta}\frac{x^\sigma}{T^{1-2\varepsilon_1}}d\sigma = O\left(\frac{x^{1+\delta}}{T^{1-2\varepsilon_1}}\right),$$

where we also used the fact that the Euler product $(1-2^{-s})^{-1/2}\prod_{p\equiv 3\,(\mathrm{mod}\,4)}(1-p^{-2s})^{-1/2}$ is absolutely bounded for $\mathrm{Re}(s) > 1/2+\varepsilon$. We get the same bound for $\int_{L_7}$.

For the vertical integral over L_2, we have

$$\int_{L_2} F(s)\frac{x^s}{s}ds \ll \int_{\eta}^{T}\frac{x^{1/2+\varepsilon}}{(t+\frac{1}{2})^{1-2\varepsilon_1}}dt = O\left(x^{1/2+\varepsilon}T^{2\varepsilon_1}\right),$$

which also holds for $\int_{L_6}$. Finally, we have

$$\int_{L_4} F(s)\frac{x^s}{s}ds \ll \eta\, x^{1+\kappa},$$

and choosing $T = x^{1/2}$ and $\eta < x^{-\frac{1}{2}-\kappa}$, this gives

$$\int_{1+\delta-iT}^{1+\delta+iT} F(s)\frac{x^s}{s}ds = \lim_{\eta\to 0^+}\left(\int_{1/2+\varepsilon-i\eta}^{1+\kappa-i\eta} - \int_{1/2+\varepsilon+i\eta}^{1+\kappa+i\eta}\right)F(s)\frac{x^s}{s}ds + O\Big(x^{1/2+\varepsilon}\Big).$$

Note that κ can be arbitrarily small, and choosing for example $\kappa = x^{-2}$, we have

$$\lim_{\eta\to 0^+}\left(\int_{1/2+\varepsilon-i\eta}^{1+\kappa-i\eta} - \int_{1/2+\varepsilon+i\eta}^{1+\kappa+i\eta}\right)F(s)\frac{x^s}{s}ds = \lim_{\eta\to 0^+}\left(\int_{1/2+\varepsilon-i\eta}^{1-i\eta} - \int_{1/2+\varepsilon+i\eta}^{1+i\eta}\right)F(s)\frac{x^s}{s}ds + O(1).$$

Putting everything together, we have

$$\sum_{n\le x} 1_{\mathbf{E}}(n) = \frac{1}{2\pi i}\int_{1/2+\varepsilon}^{1} G(\sigma)\frac{x^\sigma}{\sigma}\lim_{\eta\to 0^+}\Big((\sigma - i\eta - 1)^{-1/2} - (\sigma + i\eta - 1)^{-1/2}\Big)d\sigma + O\Big(x^{1/2+\varepsilon}\Big),$$

where $G(s) = (s-1)^{1/2}F(s)$.

We use the fact that when $\sigma \in (0,1)$, $\log(\sigma \pm i\eta - 1) \sim \log|\sigma - 1| \pm i\pi$ as $\eta \to 0^+$. Writing $(\sigma \pm i\eta - 1)^{-1/2} = \exp\left(-\frac{1}{2}\log(\sigma \pm i\eta - 1)\right)$, we see that $(\sigma \pm i\eta - 1)^{-1/2} \sim \mp i|\sigma - 1|^{-1/2}$, and we have

$$\lim_{\eta\to 0^+}\Big((\sigma - i\eta - 1)^{-1/2} - (\sigma + i\eta - 1)^{-1/2}\Big) = 2i|\sigma - 1|^{-1/2}.$$

Replacing above, this proves the theorem under the Riemann Hypothesis. Unconditionally, we start from (3.50), and we use a similar contour, but with $1/2+\varepsilon$ replaced by $1 - c/\sqrt{\log x}$, where c is small enough to insure that the contour does not contain any zeroes of $\zeta(s)$ or $L(s,\chi_4)$. Working as above, we get

$$\sum_{n\le x} 1_{\mathbf{E}}(n) = \frac{1}{\pi}\int_{1-\frac{c}{\sqrt{\log x}}}^{1}\frac{x^\sigma}{\sigma}G(\sigma)|\sigma-1|^{-1/2}d\sigma + O\left(\frac{x^{1+\delta}}{T^{1-2\varepsilon_1}} + x\exp\left(-c\sqrt{\log x}\right)T^{2\varepsilon_1}\right),$$

and choosing $\delta = 1/\log x$ and $T = \exp(c\sqrt{\log x})$, we get

$$\sum_{n\le x} 1_{\mathbf{E}}(n) = \frac{1}{\pi}\int_{1-c/\sqrt{\log x}}^{1}\frac{x^\sigma}{\sigma}G(\sigma)|\sigma-1|^{-1/2}d\sigma + O\left(x\exp(-c_0\sqrt{\log x})\right),$$

for some $c_0 > 0$. Finally, we have

$$\int_{\frac{1}{2}+\varepsilon}^{1-c/\sqrt{\log x}}\frac{x^\sigma}{\sigma}G(\sigma)|\sigma-1|^{-1/2}d\sigma \ll \int_{\frac{1}{2}+\varepsilon}^{1-c/\sqrt{\log x}}\frac{x^\sigma}{\sigma|\sigma-1|^{1/2}}d\sigma \ll x^{1-c/\sqrt{\log x}}$$

which shows the unconditional result.

3.7.2 Averages of Hardy–Littlewood constants

The following proposition is a more precise version of [Smi13, Theorem 1.1] who showed that

$$\sum_{\substack{1\le d_1,d_2\le H\\ \text{distinct}}}\mathfrak{S}(\{d_1,d_k\}) = H^2 + O(H^{1+\varepsilon}).$$

We remark that our normalization differs from [Smi13] for the singular series.

Proposition 3.7.1. *Fix $\varepsilon > 0$. There exists $c > 0$ such that*

$$\sum_{\substack{1\le d_1,d_2\le H\\ \text{distinct}}} \mathfrak{S}(\{d_1,d_k\}) = H^2+\frac{2}{\pi K^2}\int_{1/2+\varepsilon}^{1} \frac{F'(\sigma)H^\sigma + F(\sigma)H^\sigma \log H}{|\sigma-1|^{1/2}}d\sigma+O\left(H\exp(-c\sqrt{\log H})\right)$$

where $F(s) = \zeta(s-1)M(s-1)\left[(s-1)\zeta(s)\right]^{1/2} s^{-1}$, with $M(s)$ as defined by (3.31). Assuming the Riemann Hypothesis, we can replace the error term by $O\left(H^{1/2+\varepsilon}\right)$.

Proof. As in [Smi13, § 2.3], we have

$$\begin{aligned}\sum_{\substack{1\le d_1,d_2\le H\\ \text{distinct}}} \mathfrak{S}(\{d_1,d_k\}) &= 2\sum_{1\le d<H} \mathfrak{S}(\{0,d\})(H-d)\\ &= \frac{1}{K^2}\frac{1}{2i\pi}\int_{(2)}\frac{D(s)}{s(s+1)}H^{s+1}ds,\end{aligned}$$

where $D(s) = \zeta(s)\zeta(s+1)^{\frac{1}{2}}M(s)$ as defined in the beginning of section 3.5. As [Smi13], we compute the main term, coming from the pole of $D(s)$ at $s=1$, which gives

$$\sum_{\substack{1\le d_1,d_2\le H\\ \text{distinct}}} \mathfrak{S}(\{d_1,d_k\}) = H^2 + \frac{1}{K^2}\frac{1}{2i\pi}\int_{(\varepsilon)}\frac{D(s)}{s(s+1)}H^{s+1}ds.$$

We first assume the Riemann hypothesis and we evaluate the integral

$$\frac{1}{2i\pi}\int_{(\varepsilon)}\frac{D(s)}{s(s+1)}H^{s+1}ds = \frac{1}{2i\pi}\int_{(1+\varepsilon)}\frac{F(s)}{(s-1)^{3/2}}H^s ds$$

where $F(s) = \zeta(s-1)M(s-1)\left[(s-1)\zeta(s)\right]^{1/2} s^{-1}$ is analytic for $\mathrm{Re}(s) > 1/2+\varepsilon$. We begin with an integration by part to obtain

$$\begin{aligned}\int_{(1+\varepsilon)}\frac{F(s)}{(s-1)^{3/2}}H^s ds &= \lim_{T\to\infty}[-2F(s)H^s(s-1)^{-1/2}]_{1+\varepsilon-iT}^{1+\varepsilon+iT} + 2\int_{(1+\varepsilon)}\frac{F'(s)H^s + F(s)H^s\log H}{(s-1)^{1/2}}ds\\ &= 2\int_{(1+\varepsilon)}\frac{F'(s)H^s + F(s)H^s\log H}{(s-1)^{1/2}}ds.\end{aligned}$$

To evaluate the last integral, we first approximate the line integral by the segment from $1+\varepsilon-iT$ to $1+\varepsilon+iT$, and use the contour of Figure 3.2. Working as in the proof of

Theorem 3.2.1, we get

$$\begin{aligned}
&\frac{2}{2\pi i}\int_{(1+\varepsilon)} \frac{F'(s)H^s + F(s)H^s \log H}{(s-1)^{1/2}} ds \\
&= \frac{1}{\pi i}\int_{1/2+\varepsilon}^{1} \left(F'(\sigma)H^\sigma + F(\sigma)H^\sigma \log H\right) \lim_{\eta\to 0+} \left((\sigma-1-i\eta)^{-1/2} - (\sigma-1+i\eta)^{-1/2}\right) d\sigma + O\left(H^{1/2+\varepsilon}\right) \\
&= \frac{2}{\pi}\int_{1/2+\varepsilon}^{1} \left(F'(\sigma)H^\sigma + F(\sigma)H^\sigma \log H\right) |\sigma-1|^{-1/2} d\sigma + O\left(H^{1/2+\varepsilon}\right).
\end{aligned}$$

Replacing above, this gives (under the Riemann Hypothesis)

$$\frac{1}{K^2}\frac{1}{2\pi i}\int_{(\varepsilon)} \frac{D(s)}{s(s+1)} H^{s+1} ds = \frac{2}{\pi K^2}\int_{1/2+\varepsilon}^{1} \frac{F'(\sigma)H^\sigma + F(\sigma)H^\sigma \log H}{|\sigma-1|^{1/2}} d\sigma + O\left(H^{1/2+\varepsilon}\right).$$

To do a proof without the Riemann Hypothesis, we proceed as in the proof of Theorem 3.2.1, and we get

$$\frac{2}{2\pi i}\int_{(1+\varepsilon)} \frac{F'(s)H^s + F(s)H^s \log H}{(s-1)^{1/2}} ds \quad = \frac{2}{\pi}\int_{1-c_1/\sqrt{\log H}}^{1} \frac{F'(\sigma)H^\sigma + F(\sigma)H^\sigma \log H}{|\sigma-1|^{1/2}} d\sigma + O\left(H\exp\left(-c\sqrt{\log H}\right)\right).$$

To conclude the proof, we show that

$$\int_{1-c_1/\sqrt{\log H}}^{1} \frac{F'(\sigma)H^\sigma + F(\sigma)H^\sigma \log H}{|s-1|^{1/2}} d\sigma = \int_{1/2+\varepsilon}^{1} \frac{F'(\sigma)H^\sigma + F(\sigma)H^\sigma \log H}{|s-1|^{1/2}} d\sigma + O\left(H\exp\left(-c\sqrt{\log H}\right)\right).$$

This follows from the fact that ζ does not vanish on $[\frac{1}{2}+\varepsilon, 1]$, so F and F' are defined and continuous on $[\frac{1}{2}+\varepsilon, 1]$, in particular, they are uniformly bounded. We have

$$\begin{aligned}
\int_{\frac{1}{2}}^{1-c_1/\sqrt{\log H}} \frac{F'(\sigma)H^\sigma + F(\sigma)H^\sigma \log H}{|\sigma-1|^{1/2}} d\sigma &\ll_F \int_{\frac{1}{2}}^{1-c_1/\sqrt{\log H}} \frac{H^\sigma \log H}{|\sigma-1|^{1/2}} d\sigma \\
&\ll H^{1-c_1/\sqrt{\log H}} (\log H)^{\frac{5}{4}} \ll_c H\exp\left(-c\sqrt{\log H}\right)
\end{aligned}$$

for any $c < c_1$. □

We can now prove Proposition 3.1.3. We observe that it is a more precise version of (a particular case of) [FKR17, Proposition 1.3] who showed that

$$\sum_{\substack{1\le d_1,\dots,d_k\le H \\ \text{distinct}}} \mathfrak{S}(\{d_1,\dots,d_k\}) = H^k + O\left(H^{k-2/3+o(1)}\right).$$

Proof of Proposition 3.1.3. Note that the cases $k=0$ or 1 are easy. We have $\mathfrak{S}(\emptyset) = \mathfrak{S}(\{d\}) = 1$, so we obtain 1 and H respectively, without error term. The case $k=2$ is proven in

Proposition 3.7.1. Similarly to [MS04, (17)], we have

$$\sum_{\substack{1\le d_1,\dots,d_k\le H\\ \text{distinct}}} \mathfrak{S}(\{d_1,\dots,d_k\}) = \sum_{r=0}^{k} \binom{k}{r} \frac{(H-r)!}{(H-k)!} \sum_{\substack{1\le d_1,\dots,d_r\le H\\ \text{distinct}}} \mathfrak{S}_0(\{d_1,\dots,d_r\})$$
$$= \frac{H!}{(H-k)!} + \binom{k}{2} \frac{(H-2)!}{(H-k)!} \sum_{\substack{1\le d_1,d_2\le H\\ \text{distinct}}} \mathfrak{S}_0(\{d_1,d_2\}) + O(H^{k-\frac{3}{2}+\varepsilon}),$$

where we used the decomposition $\mathfrak{S}(\mathcal{H}) = \sum_{\mathcal{T}\subseteq\mathcal{H}} \mathfrak{S}_0(\mathcal{T})$, the fact that $\mathfrak{S}_0(\{d\}) = 0$, and the bound from Theorem 3.3.4 as soon as the size of the set is larger than 2. Using the estimates

$$\frac{H!}{(H-k)!} = H(H-1)\dots(H-k+1) = H^k + H^{k-1}\sum_{i=1}^{k-1}(-i) + O_k(H^{k-2})$$
$$= H^k - H^{k-1}\frac{k(k-1)}{2} + O_k(H^{k-2}),$$
$$\frac{(H-2)!}{(H-k)!} = H^{k-2} + O_k(H^{k-3}),$$

and Proposition 3.7.1, this gives

$$\sum_{\substack{1\le d_1,\dots,d_k\le H\\ \text{distinct}}} \mathfrak{S}(\{d_1,\dots,d_k\})$$
$$= H^k - \frac{k(k-1)}{2}H^{k-1} + \binom{k}{2}H^{k-2} \sum_{\substack{1\le d_1,d_2\le H\\ \text{distinct}}} (\mathfrak{S}(\{d_1,d_2\})-1) + O(H^{k-\frac{3}{2}+\varepsilon})$$
$$= H^k - \frac{k(k-1)}{2}H^{k-1} + \binom{k}{2}H^{k-2}\left(\frac{2}{\pi K^2}\int_{1/2+\varepsilon}^{1} \frac{F'(\sigma)H^\sigma + F(\sigma)H^\sigma \log H}{|\sigma-1|^{1/2}}d\sigma + H\right) + O(H^{k-\frac{3}{2}+\varepsilon})$$
$$= H^k + k(k-1)\frac{H^{k-1}}{\pi K^2}\int_{1/2+\varepsilon}^{1} \frac{F'(\sigma)H^{\sigma-1} + F(\sigma)H^{\sigma-1}\log H}{|\sigma-1|^{1/2}}d\sigma + O(H^{k-\frac{3}{2}+\varepsilon}).$$

□

3.7.3 Another formulation of Theorem 3.1.2

We conclude this section by stating a different version of Theorem 3.1.2 with a very good error term by using an integral form for the main term. We use this proposition for numerical testing in Section 3.9.

Proposition 3.7.2. *Fix* $\varepsilon > 0$ *and let* $S(q,v,H)$ *as in* (3.19). *There exists* $c > 0$ *such that for* $v \not\equiv 0 \pmod q$

$$S(q,v,H) = \frac{H}{q} + \frac{1}{2K^2\phi(q)}\sum_{\chi\neq\chi_0}\chi(v)^{-1}C_{q,\chi} + \frac{1}{2\pi K^2\phi(q)}\int_{1/2+\varepsilon}^{1}\frac{F_{\chi_0}(\sigma)H^{\sigma-1}}{|\sigma-1|^{1/2}}d\sigma + O\left(\exp(-c\sqrt{\log H})\right),$$

and

$$S(q,0,H) = \frac{H}{q} + \frac{1}{\pi K^2}\int_{1/2+\varepsilon}^{1} \frac{F'(\sigma)H^{\sigma-1} + F(\sigma)H^{\sigma-1}(\log H - A_q(\sigma)/2)}{|\sigma-1|^{1/2}} d\sigma + O\left(\exp(-c\sqrt{\log H})\right)$$

where $F(s) = \zeta(s-1)M(s-1)\left[(s-1)\zeta(s)\right]^{1/2}\Gamma(s)$ *and* $F_{\chi_0}(s) = \frac{1-q^{-(s-1)}}{s-1}F(s)$, *with* $M(s)$ *as defined by* (3.31) *and* $A_q(s) = \frac{1-q^{-(s-1)}}{s-1}$. *Assuming the Riemann Hypothesis, we can replace the error terms above by* $O\left(H^{-1/2+\varepsilon}\right)$.

Observe that close to $\frac{1}{2}$ one has $F'(\sigma) \asymp (\sigma - \frac{1}{2})^{-\frac{5}{4}}$, so the error term in the formula for $S(q,0,H)$ depends strongly on ε. The proof is similar to the other proofs of this section, and we skip the details.

Proof of Proposition 3.7.2: Starting from (3.30), we write

$$S(H) = H + \frac{1}{2K^2}\frac{1}{2\pi i}\int_{(1+\varepsilon)} \frac{F(s)}{(s-1)^{\frac{3}{2}}} H^{s-1} ds,$$

where $F(s) = \zeta(s-1)M(s-1)\left[(s-1)\zeta(s)\right]^{1/2}\Gamma(s)$, with $M(s)$ as defined by (3.31). Proceeding as in the proof of Proposition 3.7.1, with an integration by part before moving the contour of integration gives the following

$$S(H) = H + \frac{1}{\pi K^2}\int_{1/2+\varepsilon}^{1} \frac{F'(\sigma)H^{\sigma-1} + F(\sigma)H^{\sigma-1}\log H}{|\sigma-1|^{1/2}} d\sigma + O\left(\exp(-c\sqrt{\log H})\right).$$

Similarly, using (3.29) and without the integration by part, we have

$$\begin{aligned} S(H,\chi_0) &= \frac{\phi(q)}{q}H + \frac{1}{2K^2}\frac{1}{2\pi i}\int_{(1+\varepsilon)} \frac{F_{\chi_0}(s)}{(s-1)^{\frac{1}{2}}} H^{s-1} ds \\ &= \frac{\phi(q)}{q}H + \frac{1}{2\pi K^2}\int_{1/2+\varepsilon}^{1} \frac{F_{\chi_0}(\sigma)H^{\sigma-1}}{|\sigma-1|^{1/2}} d\sigma + O\left(\exp(-c\sqrt{\log H})\right), \end{aligned}$$

where

$$\begin{aligned} F_{\chi_0}(s) &= L(s-1,\chi_0)\Gamma(s-1)M_{\chi_0}(s-1)\left[(s-1)L(s,\chi_0)\right]^{1/2} \\ &= \frac{1-q^{-(s-1)}}{s-1}F(s) =: A_q(s)F(s) \end{aligned}$$

where we used $M_{\chi_0}(s) = (1-q^{-(s+1)})^{-1/2}M(s)$. Assuming the Riemann Hypothesis, we can replace the error term by $O\left(H^{-1/2+\varepsilon}\right)$. Then, we obtain the expressions in Proposition 3.7.2 by using the orthogonality of characters and expression (3.27) for the contribution of non-trivial characters as in the proof of Theorem 3.1.2. For $v \neq 0 \pmod q$, we have

$$\begin{aligned} S(q,v,H) &\sim \frac{1}{2K^2\phi(q)} \sum_{\chi \neq \chi_0} \chi(v)^{-1} C_{q,\chi} + \frac{1}{\phi(q)} S(H,\chi_0) \\ &\sim \frac{H}{q} + \frac{1}{2K^2\phi(q)} \sum_{\chi \neq \chi_0} \chi(v)^{-1} C_{q,\chi} + \frac{1}{2\pi K^2 \phi(q)} \int_{1/2+\varepsilon}^{1} \frac{F_{\chi_0}(\sigma) H^{\sigma-1}}{|\sigma-1|^{1/2}} d\sigma \end{aligned}$$

and

$$\begin{aligned} S(q,0,H) &\sim S(H) - \frac{\phi(q)}{q} H - \frac{1}{2\pi K^2} \int_{1/2+\varepsilon}^{1} \frac{F_{\chi_0}(\sigma) H^{\sigma-1}}{|\sigma-1|^{1/2}} d\sigma \\ &\sim \frac{H}{q} + \frac{1}{\pi K^2} \int_{1/2+\varepsilon}^{1} \frac{F'(\sigma) H^{\sigma-1} + F(\sigma) H^{\sigma-1} \log H}{|\sigma-1|^{1/2}} d\sigma - \frac{1}{2\pi K^2} \int_{1/2+\varepsilon}^{1} \frac{A_q(\sigma) F(\sigma) H^{\sigma-1}}{|\sigma-1|^{1/2}} d\sigma \\ &\sim \frac{H}{q} + \frac{1}{\pi K^2} \int_{1/2+\varepsilon}^{1} \frac{F'(\sigma) H^{\sigma-1} + F(\sigma) H^{\sigma-1} (\log H - A_q(\sigma)/2)}{|\sigma-1|^{1/2}} d\sigma. \end{aligned}$$

□

3.8 Heuristic in the case of r-uplets

As in [LOS16], the essence for the general conjecture in the case of the distribution of r consecutive sums of two squares is really in the particular case $r = 2$ that we explained in more details. In this section we present the heuristic for Conjecture 3.1.4 with highlights on the differences from the case $r = 2$, for this we follow again the exposition of [LOS16]. Let $r \geq 3$, $q \equiv 1 \pmod 4$ and $\mathbf{a} = (a_1, \ldots, a_r) \in \mathbb{N}^r$ be fixed. We write

$$\begin{aligned} N(x;q,\mathbf{a}) = \sum_{\substack{n \leq x \\ n \equiv a_1 \pmod q}} \sum_{\substack{h_2,\ldots,h_r > 0 \\ h_\ell \equiv a_\ell - a_{\ell-1} \pmod q}} \mathbf{1}_{\mathbf{E}}(n) \prod_{i=2}^{r} \mathbf{1}_{\mathbf{E}}(n + h_2 + \cdots + h_i) \\ \times \prod_{t=1}^{h_i - 1} (1 - \mathbf{1}_{\mathbf{E}}(n + h_2 + \cdots + h_{i-1} + t)). \end{aligned}$$

As in Section 3.4, we use the notation $\widetilde{\mathbf{1}}_{\mathbf{E}}(n) = \mathbf{1}_{\mathbf{E}}(n) - \frac{K}{\sqrt{\log n}}$, approximate all the $\log(n+t)$ by $\log x$, expand out the products and apply the Hardy–Littlewood Conjecture (3.12) in our context, neglecting the terms corresponding to products over more than 3 terms thanks to Theorem 3.3.4. Thus, heuristically, up to error of size $x(\log x)^{-\frac{r}{2}-\frac{1}{4}+\varepsilon}$, we obtain

$$N(x;q,\mathbf{a}) \sim \frac{x}{q} \Big(\frac{K}{\sqrt{\log x}} \Big)^r \alpha(x)^{-r+1} (\mathcal{D}_0(\mathbf{a},x) + \mathcal{D}_1(\mathbf{a},x) + \mathcal{D}_2(\mathbf{a},x)),$$

where $\alpha(x) = 1 - \frac{K}{\sqrt{\log x}}$ and

$$\mathcal{D}_0(\mathbf{a},x) = \sum_{\substack{h_2,\dots,h_r>0 \\ h_\ell \equiv a_\ell - a_{\ell-1} \pmod q}} \Big(1 + \sum_{1\le i<j\le r} \mathfrak{S}_0(\{0, h_{i+1}+\cdots+h_j\})\Big)\alpha(x)^{h_2+\cdots+h_r}$$

$$\mathcal{D}_1(\mathbf{a},x) = -\frac{K}{\alpha(x)\sqrt{\log x}} \sum_{\substack{h_2,\dots,h_r>0 \\ h_\ell \equiv a_\ell - a_{\ell-1} \pmod q}} \sum_{i=1}^{r}\sum_{j=2}^{r}\sum_{t=1}^{h_j-1} \mathfrak{S}_0(\{h_2+\cdots+h_i, h_2+\cdots+h_{j-1}+t\})\alpha(x)^{h_2+\cdots+h_r}$$

$$\mathcal{D}_2(\mathbf{a},x) = \frac{K^2}{\alpha(x)^2\log x} \sum_{\substack{h_2,\dots,h_r>0 \\ h_\ell \equiv a_\ell - a_{\ell-1} \pmod q}} \sum_{2\le i\le j\le r}\sum_{t_1=1}^{h_i-1}\sum_{\substack{t_2=1 \\ t_2>t_1 \text{ if } i=j}}^{h_j-1} \mathfrak{S}_0(\{t_1, h_i+\cdots+h_{j-1}+t_2\})\alpha(x)^{h_2+\cdots+h_r}.$$

Let us begin with studying $\mathcal{D}_0(\mathbf{a},x)$ in more details. As in Section 3.4, we write $H = -\frac{1}{\log\alpha(x)} \iff \alpha(x)^h = e(-h/H)$. The contribution of the 1 to $\mathcal{D}_0(\mathbf{a},x)$ gives

$$\begin{aligned}\sum_{\substack{h_2,\dots,h_r>0 \\ h_\ell \equiv a_\ell - a_{\ell-1} \pmod q}} e^{-(h_2+\cdots+h_r)/H} &= \prod_{\ell=2}^{r}\Big(\frac{H}{q} + f(a_\ell - a_{\ell-1};q) + O(H^{-1})\Big) \\ &= \Big(\frac{H}{q}\Big)^{r-1} + \Big(\frac{H}{q}\Big)^{r-2}\sum_{\ell=2}^{r} f(a_\ell - a_{\ell-1};q) + O(H^{r-3}). \quad (3.51)\end{aligned}$$

For the contribution of $\sum_{1\le i<j\le r}$ to $\mathcal{D}_0(\mathbf{a},x)$, we first make a change of variables by writing $j = i+k$, and we exchange the order of summation, which gives

$$\sum_{\substack{1\le i\le r-1 \\ 1\le k\le r-i}} \left(\sum_{\substack{h_2,\dots,h_i,h_{i+k+1},\dots,h_r>0 \\ h_\ell \equiv a_\ell - a_{\ell-1} \pmod q}} e^{-(h_2+\cdots+h_i+h_{i+k+1}+h_r)/H}\right) \times \left(\sum_{\substack{h_{i+1},\dots,h_{i+k}>0 \\ h_\ell \equiv a_\ell - a_{\ell-1} \pmod q}} \mathfrak{S}_0(\{0, h_{i+1}+\cdots+h_{i+k}\})e^{-(h_{i+1}+\cdots+h_{i+k})/H}\right).$$

For each fixed i,k, the second factor in the inner sum above is

$$\begin{aligned}&\sum_{\substack{h>0 \\ h\equiv a_{i+k}-a_i \pmod q}} \mathfrak{S}_0(\{0,h\})e^{-h/H} \sum_{\substack{h_{i+1},\dots,h_{i+k}>0 \\ h_\ell \equiv a_\ell - a_{\ell-1} \pmod q \\ h_{i+1}+\cdots+h_{i+k}=h}} 1 \\ &= \frac{1}{(k-1)!q^{k-1}} \sum_{\substack{h>0 \\ h\equiv a_{i+k}-a_i \pmod q}} \mathfrak{S}_0(\{0,h\})e^{-h/H}(h^{k-1} + O(h^{k-2})). \quad (3.52)\end{aligned}$$

We need some notation, generalizing the functions defined in Section 3.4.2. For $v,k \in \mathbb{N}$, let

$$S^{(k)}(q,v,H) := \sum_{\substack{h\geq 1 \\ h\equiv v \,(\mathrm{mod}\, q)}} \mathfrak{S}(\{0,h\})h^k e^{-h/H}$$
$$S_0^{(k)}(q,v,H) := \sum_{\substack{h\geq 1 \\ h\equiv v \,(\mathrm{mod}\, q)}} \mathfrak{S}_0(\{0,h\})h^k e^{-h/H}$$
$$S^{(k)}(H) := \sum_{h\geq 1} \mathfrak{S}(\{0,h\})h^k e^{-h/H}$$
$$S_0^{(k)}(H) := \sum_{h\geq 1} \mathfrak{S}_0(\{0,h\})h^k e^{-h/H}.$$

Note that $S^{(0)}(q,v,H) = S(q,v,H)$ as defined in (3.19). Moreover, we have

$$S_0^{(k)}(H) = S^{(k)}(H) - \sum_{h\geq 1} h^k e^{-h/H} = S^{(k)}(H) - k!H^{k+1} + O(H^{k-1})$$

and $$S_0^{(k)}(q,v,H) = S^{(k)}(q,v,H) - \sum_{\substack{h\geq 1 \\ h\equiv v \,(\mathrm{mod}\, q)}} h^k e^{-h/H} = S^{(k)}(q,v,H) - \frac{k!}{q}H^{k+1} + O(H^{k-1}).$$

Proposition 3.8.1. *Let $q \equiv 1 \,(\mathrm{mod}\, 4)$ be a prime. For any $k \geq 1$, we have*

$$S^{(k)}(H) = k!H^{k+1} - \frac{(k-1)!}{K\sqrt{\pi}} H^k (\log H)^{-\frac{1}{2}} + O(H^k (\log H)^{-\frac{3}{2}}).$$

and

$$S^{(k)}(q,v,H) = \begin{cases} \frac{k!}{q}H^{k+1} + O(H^k(\log H)^{-\frac{3}{2}}) & \text{if } v \not\equiv 0 \,(\mathrm{mod}\, q) \\ \frac{k!}{q}H^{k+1} - \frac{(k-1)!}{K\sqrt{\pi}}H^k(\log H)^{-\frac{1}{2}} + O(H^k(\log H)^{-\frac{3}{2}}) & \text{if } v \equiv 0 \,(\mathrm{mod}\, q). \end{cases}$$

We observe that the secondary term is relatively smaller in the case $k \geq 1$ than in the case $k = 0$ (which is Theorem 3.1.2). This is due to the fact that the order of the singularity is smaller when $k \geq 1$ as the poles of the functions $\zeta(k+1+s)$ and $\Gamma(s)$ do not coincide. Note also that, similarly to Theorem 3.1.2, one could develop the secondary term using a sum of descending powers of $\log H$ with explicit coefficients. We chose not to do so in this statement as we are mostly interested in the direction of the bias in the distribution of consecutive sums of two squares in arithmetic progressions.

Proof. The proof is similar to the proof of Theorem 3.1.2, and we just give a sketch. The main idea is to approximate the sums $S^{(k)}(H)$ and $S^{(k)}(q,v,H)$ via contour integration of the shifted functions $D(s-k)$ and $D_\chi(s-k)$ (for χ a character modulo q) respectively, where the functions D and D_χ are as defined in Section 3.5. For $k \geq 1$ and $\chi \neq \chi_0$, the function

$\Gamma(s)D_\chi(s-k)$ is analytic on a zero free region containing the line $\mathrm{Re}(s)=k$, thus, we have

$$\sum_{h\geq 1} 2K^2\mathfrak{S}(\{0,h\})\chi(h)h^k e^{-h/H} = O(H^k e^{-c\sqrt{\log H}}).$$

For $S^{(k)}(H)$, the function $\Gamma(s)D(s-k)$ has a simple pole at $s=k+1$ with residue $2K^2\Gamma(k+1)$ and an essential singularity at $s=k$ of the shape $(s-k)^{-\frac{1}{2}}$. We deduce that

$$\sum_{h\geq 1} 2K^2\mathfrak{S}(\{0,h\})h^k e^{-h/H} = H^{k+1}2K^2\Gamma(k+1) - 2\Gamma(k)\frac{K}{\sqrt{\pi}}H^k(\log H)^{-\frac{1}{2}} + O(H^k(\log H)^{-\frac{3}{2}}),$$

which gives

$$S^{(k)}(H) = H^{k+1}\Gamma(k+1) - \Gamma(k)\frac{1}{K\sqrt{\pi}}H^k(\log H)^{-\frac{1}{2}} + O(H^k(\log H)^{-\frac{3}{2}}).$$

In the case $\chi=\chi_0$, the function $\Gamma(s)D_{\chi_0}(s-k)$ has a simple pole at $s=k+1$ with residue $2K^2\Gamma(k+1)\frac{\phi(q)}{q}$ and an essential singularity at $s=k$ of the shape $(s-k)^{\frac{1}{2}}$. We deduce

$$\sum_{h\geq 1} 2K^2\mathfrak{S}(\{0,h\})\chi_0(h)h^k e^{-h/H} = H^{k+1}2K^2\Gamma(k+1)\frac{\phi(q)}{q} + O(H^k(\log H)^{-\frac{3}{2}}).$$

Finally, we obtain the expressions in the statement of Lemma 3.8.1 using the orthogonality relations in the case $v\not\equiv 0 \pmod q$, and the case $v\equiv 0 \pmod q$ is then deduced by subtracting the contributions of all non-zero v's to $S^{(k)}(H)$. □

Using Lemma 3.8.1 and (3.51), (3.52), we get

$$\begin{aligned}
\mathcal{D}_0(\mathbf{a},x) =& \Big(\frac{H}{q}\Big)^{r-1} + \Big(\frac{H}{q}\Big)^{r-2}\sum_{i=2}^{r} f(a_i-a_{i-1};q) \\
&+ \sum_{i=1}^{r-1}\sum_{k=1}^{r-i}\Big(\frac{H}{q}\Big)^{r-k-1}\frac{1}{(k-1)!q^{k-1}}S_0^{(k-1)}(q,a_{i+k}-a_i,H) + O(H^{r-3}) \\
=& \Big(\frac{H}{q}\Big)^{r-1} + \Big(\frac{H}{q}\Big)^{r-2}\sum_{i=1}^{r-1}\Big(S_0(q,a_{i+1}-a_i,H) + f(a_{i+1}-a_i;q)\Big) \\
&+ \Big(\frac{H}{q}\Big)^{r-2}\frac{(\log H)^{-\frac{1}{2}}}{K\sqrt{\pi}(k-1)}\sum_{\substack{1\leq i,j\leq r\\ j>i+1}}\delta(a_j\equiv a_i) + O(H^{r-2}(\log H)^{-\frac{3}{2}}).
\end{aligned}$$

Let us now study $\mathcal{D}_1(\mathbf{a},x)$. We first write

$$\mathcal{D}_1(\mathbf{a},x) = -\frac{K}{\alpha(x)\sqrt{\log x}} \sum_{\substack{h_2,\ldots,h_r>0\\ h_\ell\equiv a_\ell-a_{\ell-1}\,(\mathrm{mod}\,q)}} \Bigg(\sum_{\substack{2\le j\le r\\ 2\le i\le j}} \sum_{t=1}^{h_j-1} \mathfrak{S}_0(\{0,h_i+\cdots+h_{j-1}+t\})e^{-(h_2+\cdots+h_r)/H} + \sum_{\substack{2\le j\le r\\ j\le i\le r}} \sum_{t=1}^{h_j-1} \mathfrak{S}_0(\{h_j+\cdots+h_i,t\})e^{-(h_2+\cdots+h_r)/H} \Bigg) \tag{3.53}$$

We focus on the first inner sum of (3.53). Exchanging the order of summation, for each fixed i and $j=i+k\ge i$, we have

$$\sum_{\substack{h_i,\ldots,h_{i+k}>0\\ h_\ell\equiv a_\ell-a_{\ell-1}\,(\mathrm{mod}\,q)}} \sum_{t=1}^{h_{i+k}-1} \mathfrak{S}_0(\{0,h_i+\cdots+h_{i+k-1}+t\})e^{-(h_i+\cdots+h_{i+k})/H} \times \sum_{\substack{h_2,\ldots,h_{i-1},h_{i+k+1},\ldots,h_r>0\\ h_\ell\equiv a_\ell-a_{\ell-1}\,(\mathrm{mod}\,q)}} e^{-(h_2+\cdots+h_{i-1}+h_{i+k+1}+\cdots+h_r)/H}$$

The second sum of the above is evaluated by (3.51), and

$$\begin{aligned}
&\sum_{\substack{h_i,\ldots,h_{i+k}>0\\ h_\ell\equiv a_\ell-a_{\ell-1}\,(\mathrm{mod}\,q)}} \sum_{t=1}^{h_{i+k}-1} \mathfrak{S}_0(\{0,h_i+\cdots+h_{i+k-1}+t\})e^{-(h_i+\cdots+h_{i+k})/H}\\
&= \sum_{u>0}\mathfrak{S}_0(\{0,u\}) \sum_{\substack{h>u\\ h\equiv a_{i+k}-a_{i-1}\,(\mathrm{mod}\,q)}} e^{-h/H} \sum_{\substack{h_i,\ldots,h_{i+k-1}>0\\ h_\ell\equiv a_\ell-a_{\ell-1}\,(\mathrm{mod}\,q)\\ h_i+\cdots+h_{i+k-1}<u}} \sum_{\substack{h_{i+k}>0\\ h_{i+k}\equiv a_{i+k}-a_{i+k-1}\,(\mathrm{mod}\,q)\\ h_i+\cdots+h_{i+k}=h}} 1\\
&= \sum_{u>0}\mathfrak{S}_0(\{0,u\})e^{-u/H} \sum_{\substack{h'>0\\ h'\equiv a_{i+k}-a_{i-1}-u\,(\mathrm{mod}\,q)}} e^{-h'/H} \sum_{\substack{h_i,\ldots,h_{i+k-1}>0\\ h_\ell\equiv a_\ell-a_{\ell-1}\,(\mathrm{mod}\,q)\\ h_i+\cdots+h_{i+k-1}<u}} 1\\
&= \sum_{u>0}\mathfrak{S}_0(\{0,u\})e^{-u/H}\Big(\frac{1}{k!}(\frac{u}{q})^k + O(u^{k-1})\Big)\Big(\frac{H}{q}+O(1)\Big)\\
&= \frac{H}{k!q^{k+1}}S_0^{(k)}(H) + O(H^{k+\varepsilon})
\end{aligned}$$

We get a similar estimate for the second inner sum of (3.53) involving $\mathfrak{S}_0(\{h_j+\cdots+h_i,t\})$ by making a change of variable to replace it by $\mathfrak{S}_0(\{0,r+h_{j+1}\cdots+h_i\})$ with $r=h_j-t$, and we obtain

$$\begin{aligned}\mathcal{D}_1(\mathbf{a},x) &= -2\frac{K}{\alpha(x)\sqrt{\log x}}\sum_{i=2}^{r}\sum_{k=0}^{r-i}\Big(\frac{H}{q}\Big)^{r-2-k}\frac{H}{q^{k+1}k!}S_0^{(k)}(H) + O(H^{r-3+\varepsilon})\\ &= -2\frac{K}{\alpha(x)\sqrt{\log x}}\frac{H^{r-1}}{q^{r-1}}\Big((r-1)S_0(H) - \frac{(\log H)^{-\frac{1}{2}}}{K\sqrt{\pi}}\sum_{k=1}^{r-2}\frac{(r-1-k)}{k}\Big) + O(H^{r-2}(\log H)^{-\frac{3}{2}})\end{aligned}$$

The same ideas are used to estimate $\mathcal{D}_2(\mathbf{a},x)$. Let i and $j = i+k \geq i$ be fixed, and let us study the sum in $\mathcal{D}_2(\mathbf{a},x)$. In the case $i = j$, this is

$$\sum_{\substack{h_i>0\\ h_i\equiv a_i-a_{i-1}\,(\mathrm{mod}\,q)}}\ \sum_{1\leq t_1<t_2\leq h_i-1}\mathfrak{S}_0(\{t_1,t_2\})e^{-h_i/H} = \Big(\frac{H^2}{q}+O(H)\Big)S_0(H)$$

as we already saw in the case $r = 2$. In the case $k \geq 1$, we have

$$\begin{aligned}&\sum_{\substack{h_i,\dots,h_{i+k}>0\\ h_\ell\equiv a_\ell-a_{\ell-1}\,(\mathrm{mod}\,q)}}\ \sum_{t_1=1}^{h_i-1}\sum_{t_2=1}^{h_{i+k}-1}\mathfrak{S}_0(\{t_1,h_i+\dots+h_{i+k-1}+t_2\})e^{-(h_i+\dots+h_{i+k})/H}\\ &= \sum_{1\leq t_1<t_2'}\mathfrak{S}_0(\{0,t_2'-t_1\})\sum_{h>t_2'}e^{-h/H}\sum_{\substack{h_i,\dots,h_{i+k-1}>0\\ h_\ell\equiv a_\ell-a_{\ell-1}\,(\mathrm{mod}\,q)\\ t_1<h_i+\dots+h_{i+k-1}<t_2'}}\ \sum_{\substack{h_{i+k}>0\\ h_{i+k}\equiv a_{i+k}-a_{i+k-1}\,(\mathrm{mod}\,q)\\ h_i+\dots+h_{i+k}=h}}1\\ &= \sum_{u>0}\mathfrak{S}_0(\{0,u\})\sum_{t_2'>u}\sum_{h>t_2'}e^{-h/H}\Big(\frac{1}{k!}(\frac{u}{q})^k+O(u^{k-1})\Big)\\ &= \sum_{u>0}\mathfrak{S}_0(\{0,u\})e^{-u/H}\Big(\frac{1}{k!}(\frac{u}{q})^k+O(u^{k-1})\Big)\Big(\frac{H^2}{q}+O(H)\Big)\\ &= \frac{H^2}{k!q^{k+1}}S_0^{(k)}(H)+O(H^{k+1+\varepsilon}).\end{aligned}$$

We deduce that

$$\begin{aligned}\mathcal{D}_2(\mathbf{a},x) &= \frac{K^2}{\alpha(x)^2\log x}\sum_{i=2}^{r}\sum_{k=0}^{r-i}\Big(\frac{H}{q}\Big)^{r-2-k}\frac{H^2}{k!q^{k+1}}S_0^{(k)}(H)+O(H^{r-3+\varepsilon})\\ &= \frac{K^2}{\alpha(x)^2\log x}\frac{H^r}{q^{r-1}}\Big((r-1)S_0(H)-\frac{(\log H)^{-\frac{1}{2}}}{K\sqrt{\pi}}\sum_{k=1}^{r-2}\frac{(r-1-k)}{k}\Big)+O(H^{r-2}(\log H)^{-\frac{3}{2}}).\end{aligned}$$

Wrapping up, we obtain

$$N(x;q,\mathbf{a}) = \frac{x}{q}\Big(\frac{K}{\sqrt{\log x}}\Big)^r \alpha(x)^{-r+1}\Big(\big(\tfrac{H}{q}\big)^{r-1} + \big(\tfrac{H}{q}\big)^{r-2}\sum_{i=1}^{r-1}\big(\mathcal{D}_0(a_i,a_{i+1},x) - \tfrac{H}{q} + \mathcal{D}_1(a_i,a_{i+1},x) + \mathcal{D}_2(a_i,a_{i+1},x)\big)$$
$$- \big(\tfrac{H}{q}\big)^{r-2}\frac{(\log H)^{-\frac{1}{2}}}{K\sqrt{\pi}}\Big(\sum_{k=1}^{r-2}\sum_{i=1}^{r-1-k}\frac{\delta(a_{i+k+1}\equiv a_i) - \frac{1}{q}}{k}\Big) + O(H^{r-2}(\log H)^{-\frac{3}{2}})\Big).$$

Then using $H = \frac{\sqrt{\log x}}{K} - \frac{1}{2} + O((\log x)^{-\frac{1}{2}})$, and the estimates for $\mathcal{D}_i(a,b,x)$, $i = 0,1,2$ from Theorem 3.1.2 we obtain Conjecture 3.1.4.

3.9 Numerical data

We present in this section some numerical data testing the approximation of Conjecture 3.1.1 for $N(x;q,(a,b))$. One of the challenges of the numerical testing is the change of scale introduced by the change of variable (3.17), which gives $H = \sqrt{\log x}/K$. The actual values of $N(x;q,(a,b))$ were obtained by using SageMath [The19] on about 20 CPU cores in a Linux cluster for a couple of months, which allows us to take $x = 10^{12}$. But then, $H \approx 6.356$ in Theorem 3.1.2, which is very small even for this large value of x.

There are some technical methods for computing the Euler products, whenever they converge, and their derivatives with enough precision, and we used the following equality, which gives us a faster convergence:

$$\prod_{p\equiv 3 \pmod 4}\Big(1-p^{-2s}\Big) = \prod_{1\le j\le J}\left(\frac{L(2^j s,\chi_4)}{\zeta(2^j s)(1-2^{-2^j s})}\right)^{1/2^j}\prod_{p\equiv 3 \pmod 4}\Big(1-p^{-2^{J+1}s}\Big)^{1/2^J}.$$

Note that the rightmost hand side product converges much faster than the left hand side one. Also, its derivatives can be computed by taking the derivatives of the right hand side instead so that one might obtain some recursive formula.

We present in Table 3.5 some numerical data for Conjecture 3.1.1, for $q = 5$ and $x = 10^{12}$. There are 25 cases for $N(x;q,(a,b))$ in Table 3.5, but the conjectural asymptotic of Conjecture 3.1.1 only depends on $b-a \pmod q$, and there are then unavoidable fluctuations in the data for various pairs (a,b) with the same value of $b-a \pmod 5$. The fit between the numerical data and the conjecture is slightly better when $b-a \not\equiv 0 \pmod 5$. The numerical data is also influenced by the bias of Theorem 3.2.4, which is of smaller magnitude that the bias of Conjecture 3.1.1 but in the opposite direction, and the data when $a = b = 0$ in particular shows the influence of both biases. We have used several asymptotic approximations of our conjecture in Table 3.5. We used Conjecture 3.1.1 as such with $J = 1$ (the column labeled "Conjecture 3.1.1"), and we also used the more complicated expression of Proposition 3.4.2 for $\mathcal{D}_0(a,b;x) + \mathcal{D}_1(a,b;x) + \mathcal{D}_2(a,b;x)$ in (3.18), where we evaluate the exponential sums $E(q,v;H)$ exactly for each residue class (recall that $H = \sqrt{\log x}/K \approx 6.356$ when $x = 10^{12}$). We then replaced $S_0(q,v;H)$ in that expression by the approximation of Theorem 3.1.2 with $J = 1$ (the column labeled "Theorem 3.1.2"), and by the actual numerical value of $S_0(q,v;H)$ (the column labeled "$S_0(q,v;H)$").

We also present some numerical data for Theorem 3.1.2 in Table 3.6 and 3.7 for larger values of H. We tested the asymptotic of Theorem 3.1.2 for $J = 1{,}2{,}3$ and the integral formula of Proposition 3.7.2 for various values of H. For $H \approx 6.356$, larger values of J or the integral formula of Proposition 3.7.2 are not approximating well $S(q,v;H)$, but one can see the fit for larger values of H. The values of the constants $c_0(2), c_0(3), c_1(2), c_1(3)$ can be computed by taking more terms in the Taylor expansions of the proof of Theorem 3.1.2, similarly to the computations of $c_0(1), c(1)$ in Section 3.5. We did not include those computations (which are lengthy but straightforward and not very interesting) in the paper. The numerical values are

$$c_0(1) \approx 0.604541230, \quad c_0(2) \approx 0.696827721, \quad c_0(3) \approx 1.185903185$$
$$c_1(1) \approx -0.167588374, \quad c_1(2) \approx -0.054190676, \quad c_1(3) \approx -0.328019051.$$

a	b	$N(x;q,(a,b))$	$S_0(q,v;H)$	Theorem 3.1.2	Conjecture 3.1.1	Error1	Error2	Error3
	0	$4\ 108 \cdot 10^6$	$3\ 585 \cdot 10^6$	$3\ 219 \cdot 10^6$	$3\ 919 \cdot 10^6$	1.1461	1.2763	1.0483
	1	$7\ 153 \cdot 10^6$	$6\ 949 \cdot 10^6$	$6\ 904 \cdot 10^6$	$6\ 841 \cdot 10^6$	1.0294	1.0360	1.0457
0	2	$5\ 604 \cdot 10^6$	$5\ 430 \cdot 10^6$	$5\ 493 \cdot 10^6$	$5\ 426 \cdot 10^6$	1.0320	1.0203	1.0329
	3	$8\ 055 \cdot 10^6$	$7\ 487 \cdot 10^6$	$7\ 858 \cdot 10^6$	$7\ 153 \cdot 10^6$	1.0759	1.0250	1.1261
	4	$5\ 780 \cdot 10^6$	$5\ 626 \cdot 10^6$	$5\ 603 \cdot 10^6$	$5\ 738 \cdot 10^6$	1.0274	1.0317	1.0073
	0	$5\ 777 \cdot 10^6$	$5\ 626 \cdot 10^6$	$5\ 603 \cdot 10^6$	$5\ 738 \cdot 10^6$	1.0269	1.0312	1.0068
	1	$3\ 765 \cdot 10^6$	$3\ 585 \cdot 10^6$	$3\ 219 \cdot 10^6$	$3\ 919 \cdot 10^6$	1.0503	1.1697	0.9607
1	2	$6\ 870 \cdot 10^6$	$6\ 949 \cdot 10^6$	$6\ 904 \cdot 10^6$	$6\ 841 \cdot 10^6$	0.9886	0.9950	1.0043
	3	$5\ 354 \cdot 10^6$	$5\ 430 \cdot 10^6$	$5\ 493 \cdot 10^6$	$5\ 426 \cdot 10^6$	0.9860	0.9747	0.9868
	4	$7\ 742 \cdot 10^6$	$7\ 487 \cdot 10^6$	$7\ 858 \cdot 10^6$	$7\ 153 \cdot 10^6$	1.0341	0.9853	1.0824
	0	$8\ 050 \cdot 10^6$	$7\ 487 \cdot 10^6$	$7\ 858 \cdot 10^6$	$7\ 153 \cdot 10^6$	1.0752	1.0244	1.1254
	1	$5\ 516 \cdot 10^6$	$5\ 626 \cdot 10^6$	$5\ 603 \cdot 10^6$	$5\ 738 \cdot 10^6$	0.9804	0.9845	0.9613
2	2	$3\ 755 \cdot 10^6$	$3\ 585 \cdot 10^6$	$3\ 219 \cdot 10^6$	$3\ 919 \cdot 10^6$	1.0474	1.1664	0.9580
	3	$6\ 838 \cdot 10^6$	$6\ 949 \cdot 10^6$	$6\ 904 \cdot 10^6$	$6\ 841 \cdot 10^6$	0.9840	0.9903	0.9996
	4	$5\ 351 \cdot 10^6$	$5\ 430 \cdot 10^6$	$5\ 493 \cdot 10^6$	$5\ 426 \cdot 10^6$	0.9853	0.9741	0.9861
	0	$5\ 609 \cdot 10^6$	$5\ 430 \cdot 10^6$	$5\ 493 \cdot 10^6$	$5\ 426 \cdot 10^6$	1.0330	1.0212	1.0338
	1	$7\ 718 \cdot 10^6$	$7\ 487 \cdot 10^6$	$7\ 858 \cdot 10^6$	$7\ 153 \cdot 10^6$	1.0309	0.9822	1.0790
3	2	$5\ 549 \cdot 10^6$	$5\ 626 \cdot 10^6$	$5\ 603 \cdot 10^6$	$5\ 738 \cdot 10^6$	0.9863	0.9904	0.9670
	3	$3\ 765 \cdot 10^6$	$3\ 585 \cdot 10^6$	$3\ 219 \cdot 10^6$	$3\ 919 \cdot 10^6$	1.0503	1.1697	0.9607
	4	$6\ 867 \cdot 10^6$	$6\ 949 \cdot 10^6$	$6\ 904 \cdot 10^6$	$6\ 841 \cdot 10^6$	0.9882	0.9946	1.0039
	0	$7\ 156 \cdot 10^6$	$6\ 949 \cdot 10^6$	$6\ 904 \cdot 10^6$	$6\ 841 \cdot 10^6$	1.0298	1.0364	1.0461
	1	$5\ 357 \cdot 10^6$	$5\ 430 \cdot 10^6$	$5\ 493 \cdot 10^6$	$5\ 426 \cdot 10^6$	0.9864	0.9752	0.9872
4	2	$7\ 731 \cdot 10^6$	$7\ 487 \cdot 10^6$	$7\ 858 \cdot 10^6$	$7\ 153 \cdot 10^6$	1.0326	0.9838	1.0808
	3	$5\ 497 \cdot 10^6$	$5\ 626 \cdot 10^6$	$5\ 603 \cdot 10^6$	$5\ 738 \cdot 10^6$	0.9771	0.9812	0.9580
	4	$3\ 769 \cdot 10^6$	$3\ 585 \cdot 10^6$	$3\ 219 \cdot 10^6$	$3\ 919 \cdot 10^6$	1.0512	1.1707	0.9615

Table 3.5: The experimental value of $N(x;q,(a,b))$ versus several estimates for Conjecture 3.1.1 with $J = 1$ for $q = 5$ and $x = 10^{12}$. We used Conjecture 3.1.1 as such with $J = 1$ (the column labeled "Conjecture 3.1.1"), and we also used the more complicated expression of Proposition 3.4.2 for $\mathcal{D}_0(a,b;x) + \mathcal{D}_1(a,b;x) + \mathcal{D}_2(a,b;x)$ in (3.18), where we evaluate the exponential sums $E(q,v;H)$ exactly for each residue class (recall that $H = \sqrt{\log x}/K \approx 6.356$ when $x = 10^{12}$). We then replaced $S_0(q,v;H)$ in that expression by the approximation of Theorem 3.1.2 with $J = 1$ (the column labeled "Theorem 3.1.2"), and by the actual numerical value of $S_0(q,v;H)$ (the column labeled "$S_0(q,v;H)$"). Error1, Error2, Error3 are the percentage errors for the 4th, 5th and 6th columns, respectively.

H	$S(q,0;H) - H/q$	Prop. 3.7.2	$J = 1$	$J = 2$	$J = 3$	Prop. 3.7.2	$J = 1$	$J = 2$	$J = 3$
6.356	-0.6093	-0.0087	-0.6889	-0.4122	-0.1577	70.3362	0.8843	1.4779	3.8630
16	-0.8852	-0.5540	-1.0240	-0.8731	-0.7804	1.5980	0.8645	1.0139	1.1343
10^2	-1.3968	1.2847	-1.5059	-1.4354	-1.4094	1.0862	0.9275	0.9731	0.9910
10^4	-2.2932	-2.2839	-2.3289	-2.3040	-2.2994	1.0041	0.9846	0.9953	0.9973
10^6	-2.9169	-2.9162	-2.9337	-2.9201	-2.9184	1.0002	0.9943	0.9989	0.9995

Table 3.6: The numerical value of $S(q,0;H) - H/q$ for $q = 5$ and various values of H versus the asymptotic of Proposition 3.7.2 and Theorem 3.1.2 for $J = 1, 2,3$. The last 4 columns are the percentage errors.

H	$S(q,3;H) - H/q$	Prop. 3.7.2	$J = 1$	$J = 2$	$J = 3$	Prop. 3.7.2	$J = 1$	$J = 2$	$J = 3$
6.356	0.0327	0.0728	0.0811	0.0596	-0.0108	0.4485	0.4029	0.5485	-3.0166
16	0.0788	0.0919	0.1036	0.0919	0.0663	0.8575	0.7609	0.8581	1.1900
10^2	0.1120	0.1171	0.1262	0.1207	0.1135	0.9565	0.8875	0.9278	0.9868
10^4	0.1456	0.1461	0.1490	0.1471	0.1458	0.9966	0.9770	0.9899	0.9986
10^6	0.15813	0.15819	0.1592	0.1581	0.1577	0.9997	0.9935	1.0001	1.0030

Table 3.7: The numerical value of $S(q,3;H) - H/q$ for $q = 5$ and various values of H versus the asymptotic of Proposition 3.7.2 and Theorem 3.1.2 for $J = 1, 2,3$. The last 4 columns are the percentage errors.

Bibliography

[Ary15a] Farzad Aryan. Distribution of squares modulo a composite number. *International Mathematics Research Notices*, 2015(23):12405–12431, 2015.

[Ary15b] Farzad Aryan. The distribution of-tuples of reduced residues. *Mathematika*, 61(1):72–88, 2015.

[BBSF18] Efrat Bank, Lior Bary-Soroker, and Arno Fehm. Sums of two squares in short intervals in polynomial rings over finite fields. *American Journal of Mathematics*, 140(4):1113–1131, 2018.

[Ber12] Paul Bernays. *Über die Darstellung von positiven, ganzen zahlen durch die primitiven, binären Quadratischen formen einer nicht Quadratischen discriminante*. Dieterich, 1912.

[BFTB13] William D Banks, Tristan Freiberg, and Caroline L Turnage-Butterbaugh. Consecutive primes in tuples. *arXiv preprint arXiv:1311.7003*, 2013.

[BSF19] Lior Bary-Soroker and Arno Fehm. Correlations of sums of two squares and other arithmetic functions in function fields. *International Mathematics Research Notices*, 2019(14):4469–4515, 2019.

[BSSW16] Lior Bary-Soroker, Yotam Smilansky, and Adva Wolf. On the function field analogue of landau's theorem on sums of squares. *Finite fields and their applications*, 39:195–215, 2016.

[BW00] Antal Balog and Trevor D Wooley. Sums of two squares in short intervals. *Canadian Journal of Mathematics*, 52(4):673–694, 2000.

[CK97] RD Connors and JP Keating. Two-point spectral correlations for the square billiard. *Journal of Physics A: Mathematical and General*, 30(6):1817, 1997.

[CLM23] Alexandru Ciolan, Alessandro Languasco, and Pieter Moree. Landau and ramanujan approximations for divisor sums and coefficients of cusp forms. *Journal of Mathematical Analysis and Applications*, 519(2):126854, 2023.

[Cox13] David A. Cox. *Primes of the Form $x^2 + ny^2$*. Wiley, 2013.

[DDNS21] Chantal David, Lucile Devin, Jungbae Nam, and Jeremy Schlitt. Lemke oliver and soundararajan bias for consecutive sums of two squares. *Mathematische Annalen*, 384(3-4):1181–1242, 2021.

[EG18] Ivar Eriksson and Lukas Gustafsson. On the asymptotic behaviour of sums of two squares: Theoretical and numerical studies of the counting function b (x), 2018.

[FKR17] Tristan Freiberg, Pär Kurlberg, and Lior Rosenzweig. Poisson distribution for gaps between sums of two squares and level spacings for toral point scatterers. *arXiv preprint arXiv:1701.01157*, 2017.

[Fom98] O. M. Fomenko. Distribution of values of fourier coefficients for modular forms of weight 1. *Journal of Mathematical Sciences*, 89(1):1050–1071, Mar 1998.

[Gal76] Patrick X Gallagher. On the distribution of primes in short intervals. *Mathematika*, 23(1):4–9, 1976.

[Gol96] Elena Petrovna Golubeva. On the representation of large integers by binary quadratic forms. *Zapiski Nauchnykh Seminarov POMI*, 226:60–64, 1996.

[Gol01] EP Golubeva. Exceptional integers for binary quadratic forms. *Journal of Mathematical Sciences*, 105(4):2148–2155, 2001.

[Gor22] Ofir Gorodetsky. Sums of two squares are strongly biased towards quadratic residues, Jun 2022.

[GR21] Ofir Gorodetsky and Brad Rodgers. The variance of the number of sums of two squares in $\mathbb{F}[T]$ in short intervals. *American Journal of Mathematics*, 143(6):1703–1745, 2021.

[Gra87] Andrew Granville. Least primes in arithmetic progressions. In *Number Theory, Proc. Internat. Number Theory Conf., Laval*, pages 306–321. Citeseer, 1987.

[Hoo71] Christopher Hooley. On the intervals between numbers that are sums of two squares. 1971.

[Hoo73] C Hooley. On the intervals between numbers that are sums of two squares: Ii. *Journal of Number Theory*, 5(3):215–217, 1973.

[(ht13] Lucia (https://mathoverflow.net/users/38624/lucia). Asymptotic density of k-almost primes, 2013.

[Inc] Wolfram Research, Inc. Mathematica, Version 13.2. Champaign, IL, 2022.

[Iwa76] Henryk Iwaniec. The half dimensional sieve. *Acta Arithmetica*, 29(1):69–95, 1976.

[Kou19] Dimitris Koukoulopoulos. *The distribution of the prime numbers*, volume 203 of *Graduate Studies in Mathematics*. American mathematical society, 2019.

[Kou20] Dimitris Koukoulopoulos. *The Distribution of Prime Numbers*. Graduate Studies in Mathematics 203. American Mathematical Society, 2020.

[Lan09] Edmund Landau. *Über die Einteilung der positiven ganzen Zahlen in vier Klassen nach der Mindestzahl der zu ihrer additiven Zusammensetzung erforderlichen Quadrate*. 1909.

[LOS16] Robert J. Lemke Oliver and Kannan Soundararajan. Unexpected biases in the distribution of consecutive primes. *Proceedings of the National Academy of Sciences*, 113(31), 2016.

[Lut67] Indar Luthar. A generalization of a theorem of Landau. *Acta Arith.*, 12(3):223–228, 1967.

[Mai85] Helmut Maier. Primes in short intervals. *Michigan Mathematical Journal*, 32(2):221–225, 1985.

[May16] James Maynard. Dense clusters of primes in subsets. *Compositio Mathematica*, 152(7):1517–1554, 2016.

[MC99] Pieter Moree and Jilyana Cazaran. On a claim of ramanujan in his first letter to hardy. *Expositiones Mathematicae*, 17(4):289–311, 1999.

[Men17] Ingela Mennema. *The distribution of consecutive square-free numbers*. PhD thesis, Master Thesis, Department of Mathematics Leiden University, 2017.

[Mir49] L Mirsky. On the frequency of pairs of square-free numbers with a given difference. *Bull. Amer. Math. Soc.*, 55:936–030, 1949.

[MO06] Pieter Moree and Robert Osburn. Two-dimensional lattices with few distances. *arXiv preprint math/0604163*, 2006.

[MS04] Hugh L Montgomery and Kannan Soundararajan. Primes in short intervals. *Communications in mathematical physics*, 252(1-3):589–617, 2004.

[MV86] Hugh L Montgomery and Robert C Vaughan. On the distribution of reduced residues. *Annals of Mathematics*, 123(2):311–333, 1986.

[MV89] HL Montgomery and RC Vaughan. A basic inequality. In *Congress in Number Theory (Zarautz, 1984)*, pages 163–175. Universidad del Pai s Vasco Bilbao, 1989.

[Pra53] Karl Prachar. Über zahlen der form $a^2 + b^2$ in einer arithmetischen progression. *Mathematische Nachrichten*, 10(1-2):51–54, 1953.

[Pra56] Karl Prachar. *Primzahlverteilung*. Springer, 1956.

[RB02] Kanakanahalli Ramachandra and R Balasubramanian. Some problems of analytic number theory iv. *Hardy-Ramanujan Journal*, 25, 2002.

[Rie65] GJ Rieger. Über die anzahl der als summe von zwei quadraten darstellbaren und in einer primen restklasse gelegenen zahlen unterhalb einer positiven schranke. ii. 1965.

[RS94] Michael Rubinstein and Peter Sarnak. Chebyshev's bias. *Experimental Mathematics*, 3(3):173–197, 1994.

[Sha64] Daniel Shanks. The second-order term in the asymptotic expansion of $b(x)$. *Mathematics of Computation*, 18(85):75–86, 1964.

[Shi00] Daniel KL Shiu. Strings of congruent primes. *Journal of the London Mathematical Society*, 61(2):359–373, 2000.

[Smi13] Yotam Smilansky. Sums of two squares—pair correlation and distribution in short intervals. *International Journal of Number Theory*, 9(07):1687–1711, 2013.

[Sta28] GK Stanley. Two assertions made by ramanujan. *Journal of the London Mathematical Society*, 1(3):232–237, 1928.

[Sta29] GK Stanley. Corrigenda: two assertions made by ramanujan. *Journal of the London Mathematical Society*, pages 32–32, 1929.

[Ten15] Gérald Tenenbaum. *Introduction to analytic and probabilistic number theory*, volume 163 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2015.

[The19] The Sage Developers. *SageMath, the Sage Mathematics Software System (Version 8.6.rc0)*, 2019. `https://www.sagemath.org`.

[The22] The PARI Group, Univ. Bordeaux. *PARI/GP version* `2.13.4`, 2022. available from `http://pari.math.u-bordeaux.fr/`.

[Win13] Bruno Winckler. Th\'eor\eme de chebotarev effectif. *arXiv preprint arXiv:1311.5715*, 2013.

[Wir61] Eduard Wirsing. Das asymptotische verhalten von summen über multiplikative funktionen. *Mathematische Annalen*, 143(1):75–102, 1961.