Master thesis

On Diophantine inequalities

Universiteit Utrecht

Author:
Leah Bekkering
(4302869)

Supervised by:
Dr. Damaris Schindler
First reader:
Prof. Dr. Gunther Cornelissen
Second reader:
Dr. Valentijn Karemaker

June 18, 2020

Abstract

In this master thesis, we study Diophantine inequalities. Our aim is to solve inequalities involving polynomials with arbitrary real coefficients, where the ratio of two coefficients is irrational. In 1929, it was conjectured by Oppenheim that the inequality

$$|\alpha_1 x_1^2 + \ldots + \alpha_n x_n^2| < \delta,$$

provided $n \geq 5$, $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, not all of the same sign and such that two coefficients have irrational ratio, is soluble for all $\delta > 0$. A very interesting topic is to make this result quantitative. We study a quantitative result of Bourgain [5] on quadratic ternary diagonal forms for one parameter families and a generalisation to generic ternary diagonal forms by Schindler [33]. We generalise this previous work and study general Diophantine inequalities

$$|G_k(x_1, x_2) - \alpha_3 x_3^l| < \delta,$$

with G_k a binary form of degree $k \geq 3$ and coefficients in $\mathbb{R}$, $l \neq k$ and $\alpha_3 \in \mathbb{R}$. The goal is to find non-trivial solutions $(x_1, x_2, x_3) \in \mathbb{Z}$ where x_1, x_2 are of size N^l and x_3 of size N^k. We obtain results for the cases $G_k(x_1, x_2) = x_1^k - \alpha_2 x_2^k$ and $G_k(x_1, x_2) = x_1^k + \alpha_1 x_1^{k/2} x_2^{k/2} + \alpha_2 x_2^k$ and formulate a conjecture about a general polynomial G_k.

Acknowledgements

First of all, I would like to express my very great appreciation to Damaris Schindler for her positive contribution to my experience in mathematics. Ever since she started teaching at Utrecht University, I started to believe that I was capable of doing a mathematics master. Her determination to make the mathematics department more female-friendly and her belief in me as a mathematics student, will always stay with me. Furthermore, I would like to thank her for being my supervisor, for her comments and generous support, even though she is not working at Utrecht University anymore.
Also, I thank Gunther Cornelissen for being my first reader, Valentijn Karemaker for being my second reader, and Onno van Zomeren for his comments and suggestions on a first draft.

Contents

1 Introduction

A fundamental subject in number theory is Diophantine approximation: the approximation of irrational numbers by rational numbers. A closely related topic to approximation is solving inequalities in $\mathbb{Z}$, which are called Diophantine inequalities. In this thesis, we consider irrational ternary forms $Q(x_1, x_2, x_3)$, i.e., a ternary form having two coefficients with irrational ratio. Clearly, these forms cannot have rational solutions for $Q(\mathbf{x}) = 0$. However, we could ask ourselves if $Q(\mathbf{x})$ has solutions close to zero and how close to zero these solutions are. A well explored question is whether this is true for quadratic forms with irrational coefficients.

Let Q be a non-degenerate quadratic form over $\mathbb{R}^n$. That is, Q can be written as

$$Q(x_1, \dots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j = x^T A x, \qquad a_{ij} \in \mathbb{R},$$

where A is a symmetric real matrix $A = (a_{ij})_{1 \le i,j \le n}$, which can be transformed, after a base change, into a diagonal matrix

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

with $\lambda_i \neq 0$ for $1 \le i \le n$ (see Gerstein [14] for a broader introduction on quadratic forms).

If Q is definite, i.e., all λ_i are either positive or negative, then there exists a constant $c > 0$ such that $|Q(x)| \ge c||x||$ for all $x \in \mathbb{R}^n$, i.e., the set of values $Q(\mathbb{Z}^n)$ of Q on $\mathbb{Z}^n$ is a discrete subset of $\mathbb{R}$. Also, when Q is a rational form, which means all $a_{ij} \in \mathbb{Q}$, then clearly $Q(\mathbb{Z}^n)$ is a discrete subset of $\mathbb{R}$. A quadratic form is called irrational if it has two coefficients with an irrational ratio. In 1929, A. Oppenheim conjectured in [28] that if Q is a non-degenerate indefinite irrational quadratic form and $n \ge 3$, then $Q(\mathbb{Z}^n)$ is dense in $\mathbb{R}$. In other words,

Oppenheim Conjecture (1929) *For every $\epsilon > 0$, there exists a vector $(z_1, \dots, z_n) \in \mathbb{Z}^n \backslash \{0\}$ such that*

$$0 < |Q(z_1, \dots, z_n)| < \epsilon.$$

Partial results were proved by Davenport and Heilbronn in 1946 by using the analytical Hardy-Littlewood Circle Method [10]. They proved the conjecture for diagonal forms in five variables. A complete proof of the conjecture was given by Margulis in 1987 [24], [25], by using methods from ergodic theory.

By knowing there always exists a solution for the Diophantine inequality with arbitrary ϵ, a natural question arises: how large does a solution have to be in order to find a solution? In other words, can we make this quantitative, when do we find a solution? A quantitative result on the quadratic form $x_1 + \alpha_2 x_2^2 - \alpha_3 x_3^2$, where $\alpha_2, \alpha_3 \in \mathbb{R}\backslash\mathbb{Q}$, taken on average, is proved by Bourgain [5]. His main result is stated below.

Theorem 1.1 (Bourgain, [5]). *Consider* $Q(\mathbf{x}) = x_1^2 + \alpha_2 x_2^2 - \alpha_3 x_3^2$ *where* $\alpha_2 > 0$ *and* $\alpha_3 \in [\frac{1}{2}, 1]$. *Then, for almost all* α_3, *the following holds*

(i) Assuming the Lindelöf hypothesis for the Riemann zeta function,

$$\min_{\substack{\mathbf{x}\in\mathbb{Z}^3\backslash\{0\} \\ |\mathbf{x}|<N}} |Q(\mathbf{x})| \ll N^{-1+\epsilon} \text{ for all } \epsilon > 0.$$

Moreover, there are functions $A(N) \to \infty$ *and* $\delta(N) \to 0$ *with* $N \to \infty$ *depending on* Q, *such that*

$$\max_{|\xi|<A(N)} \min_{x\in\mathbb{Z}^3, 0<|\mathbf{x}|<N} |Q(\mathbf{x}) - \xi| < \delta(N), \tag{1.1}$$

provided

$$A(N)\delta(N)^{-2} \ll N^{1-\epsilon}.$$

(ii) Unconditionally, we have

$$\min_{\substack{\mathbf{x}\in\mathbb{Z}^3\backslash\{0\} \\ |\mathbf{x}|<N}} |Q(\mathbf{x})| \ll N^{-\frac{2}{5}+\epsilon}$$

and (1.1) *holds, assuming*

$$A(N)^3\delta(N)^{-\frac{11}{2}} \ll N^{1-\epsilon}.$$

In the proof, similar methods to the article of Blomer, Bourgain, Radziwiłł and Rudnick [2] are used. An earlier result on the quantitative Oppenheim Theorem is given by Lindenstraus and Margulis in [23], where $A(N)$ and $\delta(N)$ as defined in Theorem 1.1, are depending logarithmically on N. Another quantitative result on the Oppenheim Conjecture to mention is from Ghosh and Kelmer, where the inequality in Theorem 1.1(i) is established for generic members in the family of all indefinite ternary quadratic forms. This family is 5-dimensional, while in Theorem 1.1 a one-dimensional family is considered. Generally, a one-dimensional family is considered harder, as there is less to average over.

Later on, Schindler [33] generalised Bourgain's result to ternary forms of degree $k \geq 3$. She uses the idea of translating a Diophantine inequality into counting rational points on a planar curve, where results of Huang [18] are being used. For

$$P(\mathbf{x}) = x_1^k - \alpha_2 x_2^k - \alpha_3 x_3^k,$$

the bounds of Theorem 1.1 can be replaced with

$$N^{k-3+\epsilon}$$

assuming the Lindelöf hypothesis for the Riemann zeta function, and

$$N^{k-12/5+\epsilon}$$

for the unconditional case.

One could ask if this problem can be generalised even further. In this master thesis, we generalise the results of Schindler and Bourgain on several forms. We take a similar approach, but explain the advanced techniques and ideas in the proof more thoroughly than is done in [5]. After reading this thesis, the reader is able to understand the steps taken in [5] and [33].

The first natural generalisation, for which a quantitative version is found, is for the ternary form
$$x_1^k - \alpha_2 x_2^k - \alpha_3 x_3^l,$$
with α_2, α_3 having similar conditions as in Theorem 1.1 and $l, k \in \mathbb{Z}$, $l \geq k$ fixed. The result is stated in the theorem below.

Theorem 1.2. *Let $\alpha_2 > 0$ and $k, l \in \mathbb{Z}$, with $l \geq k$ be fixed. Then for almost all $\alpha_3 \in [\frac{1}{2}, 1]$, the following holds.*

(i) Assuming the Lindelöf hypothesis for the Riemann zeta function,
$$\min_{\substack{\mathbf{x} \in \mathbb{Z}^3 \\ |x_1|, |x_2| \sim N^l, |x_3| \sim N^k}} |x_1^k - \alpha_2 x_2^k - \alpha_3 x_3^l| \ll N^{kl-2l-k+\epsilon},$$
for any $\epsilon > 0$, where the constant depends on α_2, α_3 and ϵ.

(ii) Unconditionally, one has
$$\min_{\substack{\mathbf{x} \in \mathbb{Z}^3 \\ |x_1|, |x_2| \sim N^l, |x_3| \sim N^k}} |x_1^k - \alpha_2 x_2^k - \alpha_3 x_3^l| \ll N^{kl-\frac{12}{5}k+\epsilon},$$
for any $\epsilon > 0$. Here the constant depends on α_2, α_3 and ϵ as well.

In the proof of this theorem, we need some general theorems from Fourier analysis and the Riemann zeta function. These concepts are introduced in Sections 3 and 5. Also, some theorems in measure theory are being used. These are introduced in Section 6.1.

The bound in Theorem 1.2 *(i)* is essentially optimal. For $|x_1|, |x_2| \sim N^l$ and $|x_3| \sim N^k$, assuming values of $x_1^k - \alpha_2 x_2^k - \alpha_3 x_3^l$ are uniformly distributed, we expect to find $\sim \delta N^{2l+k-kl}$ solutions to
$$|x_1^k - \alpha_2 x_2^k - \alpha_3 x_3^l| < \delta. \tag{1.2}$$
Therefore, we expect at least one solution for (1.2) when $\delta \gg N^{kl-2l-k}$.

One can also consider the inhomogeneous case, i.e.,
$$|x_1^k - \alpha_2 x_2^k - \alpha_3 x_3^l - \xi| < \delta,$$
where $|\xi| \leq N^{kl}$ is a fixed real parameter. Here the same heuristic holds for δ. See also Figure 1 in Section 9 for a visualisation of a similar heuristic, where we consider the distance between two solutions.

Furthermore, we look at the Diophantine inequality

$$|x_1^k - g(x_2) - \alpha_3 x_3^l| < \delta,$$

where $g(x_2) := \alpha_{2,0} + \alpha_{2,1}x_2 + \ldots + \alpha_{2,k}x_2^k$ is a polynomial of degree k with $\alpha_{2,k}$ an irrational coefficient. It turns out such Diophantine inequalities have a lot of similarities with Theorem 1.2. This inequality is explored in Section 9.2.

We can ask ourselves what would happen if we generalise Theorem 1.2 even more, by finding results for

$$|G_k(\mathbf{x}) - \alpha_3 x_3^l| < \delta, \tag{1.3}$$

where $G_k(\mathbf{x})$ is a binary form of degree k, defined as

$$G_k(\mathbf{x}) := x_1^k + \sum_{i=1}^{k} \alpha_{2,i} x_1^i x_2^{k-i}, \tag{1.4}$$

with $\alpha_{2,1}, \ldots \alpha_{2,k} \in \mathbb{R}$. We explore a quadratic version in Section 9.3 which gives us the following result.

Theorem 1.3. *Let k be even and $G_k(\mathbf{x}) = x_1^k + \alpha_1 x_1^{\frac{k}{2}} x_2^{\frac{k}{2}} + \alpha_2 x_2^k$. Let $k, l \in \mathbb{Z}$, $l \geq k$ fixed. Furthermore, let $\alpha_1, \alpha_2 > 0$ such that the function*

$$h(z_3, z_4) := (-\tfrac{1}{2}\alpha_1 + \tfrac{1}{2}(\alpha_1^2 - 4\alpha_2 + 4G_k(z_3, z_4))^{\frac{2}{k}}$$

satisfies

$$C_1 < |\det \nabla^2 h(z_3, z_4)| < C_2$$

for all $(z_3, z_4) \in \mathcal{D}$, where $\mathcal{D} \subset \mathbb{R}^2$ is a connected open bounded set, C_1, C_2 positive constants and $\nabla^2 h(z_3, z_4)$ the Hessian of $h(z_3, z_4)$.
Then for almost all $\alpha_3 \in [\frac{1}{2}, 1]$, replacing $x_1^k - \alpha_2 x_2^k$ with $G_k(\mathbf{x})$ in Theorem 1.2 leads to the same results.

One could see Theorem 1.2 as a special case of Theorem 1.3. However, note that Theorem 1.2 gives a stronger result than Theorem 1.3, as we do not need to assume conditions on the Hessian of a function defined as in Theorem 1.3. In other words, we are able to prove that the condition on the Hessian holds for $x_1^k - \alpha_2 x_2^k$, so an assumption is not necessary. Therefore, we state the two cases as different theorems.

The results proved in Section 9 show us that, considering the Diophantine inequalities (1.3), the most important condition for a general polynomial $G_k(\mathbf{x})$ is its size, i.e., the size of the variables x_1, x_2, x_3. The following conjecture is motivated by this observation.

Conjecture 1.4. *Let $G_k(\mathbf{x})$ be as in (1.4). Let $\alpha_{2,1}, \ldots, \alpha_{2,k} > 0$ be such that, considering the difference $G_k(y_1, y_2) - G_k(y_3, y_4)$, we can write $y_1 = \Phi(y_2, y_3, y_4)$, where Φ in its turn can be written as $y_i\tilde{\Phi}(\mathbf{z})$ for $i = 2, 3$ or 4, such that $\tilde{\Phi}$ is a smooth function on a domain $\mathcal{D}$, with*

$$C_1 < |\det \nabla^2 \tilde{\Phi}(\mathbf{z})| < C_2$$

for all $\mathbf{z} \in \mathcal{D}$*. Then for almost all* $\alpha_3 \in [\frac{1}{2}, 1]$*,*

$$\min_{\substack{\mathbf{x}\in\mathbb{Z}^3 \\ |x_1|,|x_2|\sim N^l,|x_3|\sim N^k}} |G_k(x_1, x_2) - \alpha_3 x_3^l| \ll N^{kl-2l-k+\epsilon},$$

for any $\epsilon > 0$*, where the constant depends on* $\alpha_{2,1}, \ldots, \alpha_{2,k}, \alpha_3$ *and* ϵ*.*

If we assume the values of $G_k(x_1, x_2) - \alpha_3 x_3^l$ to be uniformly distributed, we would indeed expect to find $\sim \delta N^{2l+k-kl}$ solutions to (1.3) for $|x_1|, |x_2| \sim N^l$ and $|x_3| \sim N^k$.

The idea behind the notion of $\Phi(y_2, y_3, y_4)$ is motivated by the Implicit Function Theorem. We give a proof of this theorem in Section 7. In Section 9.4, a motivation and explanation for the conjecture is given.

When looking at quantitative results for Diophantine inequalities, one can also consider other properties than smallest solutions. If we consider more solutions, then looking at the gaps between these values gives us some interesting insights. In [2], Blomer, Bourgain, Radziwiłł and Rudnick considered generic diagonal forms $\alpha m^2 + n^2$, $\alpha > 0$ and approached the problem of the smallest gap between two such values. As mentioned before, the methods in the proof are similar to [5]. In [32], a similar problem is considered or generic binary quadratic forms $\alpha m^2 + mn + \beta n^2$. An alternative approach to this problem is given in [4]. Other results on the distribution of generic quadratic forms are obtained in [13], but these results are not quantitative.

Results on smallest gaps between values of polynomials gives us information about the distribution values of a Diophantine inequality. We introduce some well-known theorems on distribution in Section 6.2 and consider gaps between values of $\alpha m^2 + n^2 \leq X$ where $\alpha \in \mathbb{R}\backslash\mathbb{Q}$, $m, n \in \mathbb{Z}$, in Section 10. The theorem we prove is stated below.

Let λ_j be the eigenvalues of the Dirichlet Laplacian of a rectangular billiard with width $\pi/\sqrt{\alpha}$ and height π. Then each λ_j is of the form $\alpha m^2 + n^2$ with integers $m, n \geq 1$. Therefore, let

$$\#\{j : \lambda_j \leq X\} = \#\{(m, n) : m, n \geq 1, \alpha m^2 + n^2 \leq X\}.$$

The size of the smallest gap between two λ_i's is defined by

$$\delta_{\min}^{(\alpha)}(N) = \min(\lambda_{i+1} - \lambda_i : 1 \leq i < N).$$

Theorem 1.5 (Theorem 1.2 in [2])**.** *For almost all* $\alpha > 0$ *in the sense of Lebesgue measure, we have*

$$\delta_{\min}^{(\alpha)}(N) \ll \frac{1}{N^{1-\epsilon}}$$

for any $\epsilon > 0$ *and all* N*.*

We follow the steps of [2], but give a more elaborate proof in which details of the steps are more worked out. Furthermore, we mention some ideas about a general case, i.e., we look at the Diophantine inequality $\alpha m^k + n^2 \leq X$, $k \geq 3$, and repeat some steps of the quadratic case.

Lastly, in Section 8 and 11, other techniques and results in analytic number theory and Diophantine approximation are given. In Section 8, we introduce the Dimension Growth Conjecture and some results of Huang [17],[18] which are used in the proof of Theorem 1.2 and Theorem 1.3. In addition, we discuss the limitations of these results and the relation to the Dimension Growth Conjecture. In Section 11, we discuss the well-known Hardy-Littlewood Circle Method and explain some similarities and differences of this method compared to the method in Section 9.

2 Notation

We write $\mathbf{x} \in \mathbb{R}^n$, for a point $\mathbf{x} = (x_1, \dots, x_n)$ where $x_i \in \mathbb{R}$ for $1 \leq i \leq n$. We use the notation $||x||$ for $x \in \mathbb{R}$ for the minimal distance of x to the nearest integer. That is,

$$||x|| := \min_{n \in \mathbb{Z}} |x - n|.$$

We write $\{x\}$ for the fractional part of a real number x. Furthermore, we use the Vinogradov notations $\ll$, O and o. The implied constants are independent of α_3, N and δ unless stated otherwise. They may depend on α_2. When $f(x) \ll g(x)$ and $g(x) \ll f(x)$, we write $f(x) \asymp g(x)$. We write $e(z) := e^{2\pi i z}$ and $e_q(z) := e(\frac{z}{q})$. We denote the punctured disc with center $z_0 \in \mathbb{C}$ and radius $r > 0$ by

$$D^0(z_0, r) := \{z \in \mathbb{C} : 0 < |z - z_0| < r\}.$$

3 Harmonic Analysis

The techniques that are being used in the proofs in Section 9, are concerned with Fourier and Mellin transforms. In this section, we cover the basic material of these two transforms that is needed to understand the steps that are being taken in the proofs.

3.1 Fourier Analysis

The theorems and proofs of this section can be found in most of the general analysis books, see for example the book on harmonic analysis of Stein [39]. We start off by introducing Fourier coefficients and their convergence.

Theorem 3.1. *Let f be a piecewise C^1 function on $\mathbb{R}$ which is T-periodic, i.e., $f(x+T) = f(x)$ for all $x \in \mathbb{R}$. Define the Fourier coefficients of f by the formula*

$$c_n(f) = \frac{1}{T}\int_0^T f(t)e^{-2\pi i n t/T}dt.$$

Furthermore, define

$$f(x^+) = \lim_{\substack{z\to x\\ z>x}} f(z), \qquad f(x^-) = \lim_{\substack{z\to x\\ z<x}} f(z).$$

Then f admits a Fourier series $\sum_{n\in\mathbb{Z}} c_n(f)e^{2\pi i n x/T}$, that converges for all x. We have

$$\frac{f(x^+)+f(x^-)}{2} = \sum_{n\in\mathbb{Z}} c_n(f)e^{2\pi i n x/T}.$$

In addition, if f is C^1 everywhere, then its Fourier series converges uniformly and absolutely pointwise to f.

We say f is in $L^1(\mathbb{R})$ if $\int_{\mathbb{R}} |f(t)|dt$ converges.

Definition 3.2. *If $f \in L^1(\mathbb{R})$, we define its Fourier transform $\hat{f}$ by*

$$\hat{f}(x) = \int_{\mathbb{R}} f(t)e^{-2\pi i x t}dt.$$

Sometimes, a different normalisation is used. This results in a different exponential and a normalisation factor $\frac{1}{2\pi i}$ in front of the integral. Therefore, to avoid confusion, in most sources the Fourier transform is explicitly defined.
An often used identity of Fourier transform is the following.

Proposition 3.3. *(Parseval-Bessel) If f and g are piecewise continuous and T-periodic on $\mathbb{R}$ and $c_n(f), c_n(g)$ are their Fourier coefficients, we have*

$$\frac{1}{T}\int_0^T f(t)\overline{g(t)}dt = \sum_{n\in\mathbb{Z}} c_n(f)\overline{c_n(g)}.$$

In particular, we have $\frac{1}{T}\int_0^T |f(t)|^2 dt = \sum_{n\in\mathbb{Z}} |c_n(f)|^2$ and the series on the right-hand side are convergent.

We continue this section by giving some very useful properties of the Fourier transform. Again, we state these theorems without proofs as they can be found in [39] and many other analysis books.

Theorem 3.4. *(Inverse Fourier transform) If both f and $\hat{f}$ are in $L^1(\mathbb{R})$, then we have for all x where f is continuous*

$$f(x) = \int_{\mathbb{R}} \hat{f}(t) e^{2\pi i x t} dt.$$

The next identity is a very helpful tool to switch between the expression for a function and its Fourier transform.

Theorem 3.5. *(Poisson summation) Assume that f is a continuous function on $\mathbb{R}$ and that $f \in L^1(\mathbb{R})$. Then*

$$\sum_{n\in\mathbb{Z}} f(x+n) = \sum_{m\in\mathbb{Z}} \hat{f}(m) e^{2\pi i m x},$$

if both sides converge absolutely and uniformly. In particular, $\sum_{n\in\mathbb{Z}} f(n) = \sum_{m\in\mathbb{Z}} \hat{f}(m)$.

We will make use of this theorem in the proofs in Section 9 and Section 10.

A similar statement, in which the relation between a function and its Fourier transform is given is Parseval's identity.

Theorem 3.6. *(Parseval's identity) Let f be a function on $\mathbb{R}$ and $\hat{f}$ its Fourier transform. Then*

$$\int_{\mathbb{R}} |f(x)|^2 dx = \int_{\mathbb{R}} |\hat{f}(t)|^2 dt.$$

Lastly, we give two well-known examples of Fourier transforms.

Example 3.7. Let $g, h \in L^1(\mathbb{R})$.

(i) Let $g(x) = e^{-\pi x^2}$. Then $\hat{g}(y) = g(y)$.

(ii) For any nonzero $a \in \mathbb{R}$, we have $\hat{h}(x/\sqrt{a}) = \sqrt{a} h(x/\sqrt{a})$.

3.2 Mellin transform

The Mellin transform of f is defined by

$$\mathcal{M}(f)(s) = \int_0^\infty f(t) t^{s-1} dt.$$

Mellin transformation is a basic tool to analyse the behaviour of many important functions, such as the Riemann zeta function, which will be introduced in Section 5. We build up enough knowledge to understand Theorem 3.10, which is of importance in the proof of Theorem 1.2. The Mellin transform is a version of the Fourier transform in the following way [8].

Proposition 3.8. *Assume that f is continuous on $(0, \infty)$, that $f(t) = O(t^{-\alpha})$ for some $\alpha \in \mathbb{R}$ as $t \to 0$, and that $f(t)$ tends to 0 faster than any power of t as $t \to \infty$. Then the following holds:*

(a) The Mellin transform of f converges absolutely for Re $s > \alpha$ and defines a holomorphic function in that right half-plane.

(b) If we let $s = \sigma + iT$ with $\sigma > \alpha$ and set $g_\sigma(t) = e^{-2\pi\sigma t} f(e^{-2\pi t})$, then

$$\mathcal{M}(f)(s) = 2\pi \hat{g}_\sigma(T).$$

(c) We have the Mellin inversion formula which is valid for $\sigma > \alpha$:

$$f(x) = \frac{1}{2\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} x^{-s} \mathcal{M}(f)(s) ds$$

for all $x > 0$.

Let α be a complex number and $\lambda \in \mathbb{R}_{>0}$. Then the Mellin transform satisfies the following modifications (for more modifications see also the appendix of D. Zagier in [47]).

Function	Mellin transform
$f(t)$	$\mathcal{M}(f)(s)$
$f(\lambda t)$	$\lambda^{-s}\mathcal{M}(f)(s)$
$t^\alpha f(t)$	$\mathcal{M}(f)(s+\alpha)$
$f'(t)$	$(1-s)\mathcal{M}(f)(s-1)$

We give some examples of Mellin transforms which involve the Gamma function. The Gamma function is defined by

$$\Gamma(s) := \int_0^\infty x^{s-1} e^{-x} dx.$$

We have

$$\mathcal{M}(e^{-x})(s) = \Gamma(s).$$

We can deduce some important properties of the Gamma function from its Mellin transform. For example, we can deduce that $\Gamma(s)$ has a simple pole at $s = 0$ and that $\Gamma(s+1) = s\Gamma(s)$. More on poles can be found in Section 4. Furthermore, we have

$$\Gamma(n+1) = n!.$$

Two other examples of Mellin transforms are

$$\mathcal{M}\left(\frac{1}{1+t}\right)(s) = \frac{\pi}{\sin(\pi s)}$$

and

$$\mathcal{M}(e^t - 1)^{-1} = \Gamma(s)\zeta(s).$$

The Mellin transform is useful in the study of Dirichlet series. Recall that a Dirichlet series is of the form

$$\xi(s) = \sum_{n=1}^\infty a_n n^{-s},$$

where $s \in \mathbb{C}$. A Dirichlet series has an abscissa of convergence $\sigma_c \in \mathbb{C}$, such that the series converges for all $s \in \mathbb{C}$ with Re $s > \sigma_c$. For example, the Riemann zeta function, defined in Section 5, has abscissa of convergence 1.

According to Montgomery and Vaughan in [26], if $A(x) = \sum_{n \leq x} a_n$, then

$$\sum_{n=1}^{\infty} a_n n^{-s} = s \int_1^{\infty} A(x) x^{-s-1} dx$$

is also a Mellin transform. Note that the integral is from 1 to ∞ here, since $A(x)$ is defined from $x = 1$ on. We can also define the inverse Mellin transform for the Dirichlet series. This is done by Perron's formula, which is stated below in the truncated version, i.e., we define the integral for T, and let $T \to \infty$ if possible.

Lemma 3.9 (Perron's formula)**.** *If $\sigma_0 > \max(0, \sigma_c)$ and $x > 0$, where σ_c is the abscissa of convergence, then*

$$\sum_{n \leq x}{}' a_n = \frac{1}{2\pi i} \int_{\sigma_0 - iT}^{\sigma_0 + iT} \xi(s) \frac{x^s}{s} ds,$$

where $\sum'$ indicates that if x is an integer, then the last term is counted with weight 1/2.

Perron's formula is also an important tool in the proof of the Prime Number Theorem, see for example the proof in [26].

Let $T \to \infty$. The inverse Mellin transform for Dirichlet series can now be given by

$$A(x) = \frac{1}{2\pi i} \int_{\sigma_0 - i\infty}^{\sigma_0 + i\infty} \xi(s) \frac{x^s}{s} ds,$$

where $\sigma_0 > \max(\sigma_c, 0)$ with σ_c the abscissa of convergence of $\xi(s)$. We finish this section by giving a more general theorem concerning the Mellin transform, which will be used in the proofs of Theorem 1.2 and 1.5.

Theorem 3.10. *Let $\xi(s) = \sum_{n \geq 1} a_n n^{-s}$ be a Dirichlet series and $A(x) = \sum_{n \leq x} a_n$. Let $\phi(x)$ be a function with Mellin transform $\check{\phi}(x) := \mathcal{M}(\phi)(s)$. Then*

$$\sum_{n \geq 1} a_n \phi\left(\frac{n}{N}\right) = \frac{1}{2\pi i} \int_{c - i\infty}^{c + i\infty} \xi(s) \check{\phi}(s) N^s ds,$$

where the Dirichlet series converges absolutely for Re$(s) = c$.

This theorem can be proved by using the Mellin convolution theorem as in [46] or other analytic number theory books.

4 Complex analysis

In this introduction on complex analysis, we give some important definitions and theorems that will be used in the proofs of the theorems introduced in Section 1. More explicitly, it turns out in equation (9.21), that the function $F_2(t)$ defined in equation (9.6) can be written as

$$F_2(t) = \frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty} \zeta(s-it)\breve{\omega}_3(s)N^{ls}ds,$$

with $\zeta(s)$ the Riemann zeta function and $\breve{\omega}_3$ the Mellin transform of a certain smooth weight function ω_3 that is defined in Section 9. For the definition of $\zeta(s)$ and the Mellin transform we refer to Section 5 and Section 3 respectively. In this section, we will show how complex analysis can be used to calculate the integral of $F_2(t)$. Proofs of the stated theorems can be found in any complex analysis book, for example in *Complex Analysis* from S. Lang [21]. We will cover some material on calculating contour integrals. This includes holomorphic and meromorphic functions, poles, Cauchy's integral formula and the Residue theorem. We start with defining path integrals.

Let $g : [a, b] \to \mathbb{C}$ be a continuous function with $a, b \in \mathbb{R}$, $a < b$. We call two functions $g_1 : [a, b] \to \mathbb{C}$, $g_2 : [c, d] \to \mathbb{C}$ equivalent if there is a continuous monotone increasing function $\phi : [a, b] \to [c, d]$ with $g_1 = g_2 \circ \phi$. We call such equivalence classes *paths* in $\mathbb{C}$.
A function $g : [a, b] \to \mathbb{C}$ representing a path γ is called a parametrisation of the path. A continuously differentiable path is a path represented by a continuously differentiable function $g : [a, b] \to \mathbb{C}$. We define $g(a)$ as start point and $g(b)$ end point.
A path γ is called *closed* if its start point and end point are equal to each other. A closed path that has no self-intersections and is traversed counterclockwise is called a *contour*.

If γ_1, γ_2 are two paths such that the end point of γ_1 is equal to the start point of γ_2, then $\gamma_1 + \gamma_2$ is defined to be the path obtained by first traversing γ_1 and then γ_2. If γ is a path, then $-\gamma$ is the path traversed in the opposite direction.

Let γ be a continuously differentiable path and $f : \gamma \to \mathbb{C}$ a continuous function. If $g : [a, b] \to \mathbb{C}$ is a continuously differentiable parametrisation of γ, then

$$\int_\gamma f := \int_a^b f(g(t))g'(t)dt.$$

We define the *length* of γ by

$$L(\gamma) := \int_a^b |g'(t)|dt.$$

Furthermore, if $\gamma = \gamma_1 + \ldots + \gamma_r$ is a piecewise continuously differentiable path with continuously differentiable pieces γ_i for $i = 1, \ldots, r$, then for $f : \gamma \to \mathbb{C}$ a continuous function, we have

$$\int_\gamma f := \sum_{i=1}^r \int_{\gamma_i} f$$

and

$$L(\gamma) := \sum_{i=1}^{r} L(\gamma_i).$$

From now on, a path is assumed to be piecewise continuously differentiable, unless stated otherwise. A function $F : U \to \mathbb{C}$ on an open subset $U \subset \mathbb{C}$ is called analytic if for every $z \in U$ the limit

$$F'(z) = \lim_{h \in \mathbb{C}, h \to 0} \frac{F(z+h) - F(z)}{h}$$

exists. If γ is a path with start point z_0 and end point z_1, then

$$\int_\gamma F' = F(z_1) - F(z_0).$$

Definition 4.1. *Let $U \in \mathbb{C}$ and γ_1, γ_2 be two paths in U with the same start point z_0 and end point z_1. Then γ_1, γ_2 are called homotopic in U if γ_1 can be continuously be deformed into the other within U.*
In other words, there exist parametrisations $f : [0,1] \to \mathbb{C}$ of γ_1 and $g : [0,1] \to \mathbb{C}$ of γ_2, and a continuous map $H : [0,1] \times [0,1] \to U$, such that

$$\begin{aligned} H(0,t) &= f(t), \quad H(1,t) = g(t) \qquad \text{for } 0 \le t \le 1, \text{ and} \\ H(s,0) &= z_0, \qquad H(s,1) = z_1 \qquad \text{for } 0 \le s \le 1. \end{aligned}$$

Before introducing some important theorems of complex analysis, we need the definition of an analytic function.

Definition 4.2. *Let U be a non-empty open subset of $\mathbb{C}$ and $f : U \to \mathbb{C}$ a function. We call f analytic or holomorphic in $z_0 \in U$ if*

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

exists. If the limit exists, we denote it by $f'(z_0)$. If f is analytic in every $z \in U$, we call f analytic.

Remark 4.3. The definitions holomorphic and analytic will be used simultaneously, although analytic functions are primary defined in the topic of convergent power series, whereas holomorphism is originally the definition for a function that only has zeroes and no poles or other singularities. Later, it is proved that a function is analytic if and only if it is holomorphic.

Lastly, f is called analytic around z_0 if there exists some open disc $D(z_0, \delta)$, $\delta > 0$ in which f is analytic. For two analytic functions on an open subset $U \subset \mathbb{C}$, the usual rules for differentiation hold.

A very important theorem in complex analysis is the following theorem from Cauchy.

Theorem 4.4. *Let U be a non-empty open subset of $\mathbb{C}$ and $f : U \to \mathbb{C}$ an analytic function. Let γ_1, γ_2 be two homotopic paths in U with the same start point and end point. Then*

$$\int_{\gamma_1} f = \int_{\gamma_2} f.$$

This leads to the following corollary, which is often used in analytic number theory.

Corollary 4.5. *Let U be a non-empty, open, simply connected subset of $\mathbb{C}$ and $f : U \to \mathbb{C}$ an analytic function. Then for any closed path γ in U,*

$$\oint_\gamma f = 0.$$

We now state Cauchy's integral formula.

Theorem 4.6. *Let γ be a contour in $\mathbb{C}$ and U an open subset of $\mathbb{C}$ containing γ and its interior. Let z_0 be a point in the interior of γ and $f : U \to \mathbb{C}$ an analytic function. Then*

$$\frac{1}{2\pi i} \oint_\gamma \frac{f(z)}{z - z_0} dz = f(z_0).$$

If f is an analytic function on the punctured disc $D^0(z_0, r)$, $r > 0$, then z_0 is called an *isolated singularity* of f. If there exists an analytic function g on the non-punctured disc $D(z_0, r)$, such that $g(z) = f(z)$ for $z \in D^0(z_0, r)$, then z_0 is called a removable singularity of f. These definitions are important when looking at Laurent series expansions.

Theorem 4.7. *Let U be a non-empty open subset of $\mathbb{C}$ and $f : U \to \mathbb{C}$ an analytic function. Let $z_0 \in U$ and $R > 0$ such that $D^0(z_0, R) \subset U$. Then f has a Laurent series expansion*

$$f(z) = \sum_{n=-\infty}^{\infty} a_n (z - z_0)^n$$

that is convergent for $z \in D^0(z_0, R)$. For $n \in \mathbb{Z}$ we have

$$a_n = \frac{1}{2\pi i} \oint_{\gamma_{z_0, r}} \frac{f(z)}{(z - z_0)^{n+1}} dz$$

for any $0 < r < R$.

We say f has a Laurent expansion $\sum a_n (z - z_0)^n$ around z_0 if there exists a $r > 0$ such that f is equal to this Laurent expansion on $D^0(z_0, r)$. Note that z_0 is a removable singularity for f if $a_n = 0$ for all $n < 0$.

Definition 4.8. *For f with a Laurent expansion $\sum_{n=-\infty}^{\infty} a_n (z - z_0)^n$ around z_0, we define the order of f at z_0 by*

$$\mathrm{ord}_{z_0}(f) := \inf\{\textit{all } k \in \mathbb{Z} \textit{ such that } a_k \neq 0\}.$$

We see that f is holomorphic at z_0 if and only if $\mathrm{ord}_{z_0}(f) \geq 0$. Furthermore, we call the point z_0 an essential singularity of f if $\mathrm{ord}_{z_0}(f) = -\infty$, a pole or order k if $k > 0$ and $\mathrm{ord}_{z_0}(f) = -k$, and a zero of order k if $k > 0$ and $\mathrm{ord}_{z_0}(f) = k$. If the pole or zero has order 1, we call it a simple pole respectively a simple zero.

A complex function f is called meromorphic around z_0 if f is analytic on $D^0(z_0, r)$ for a $r > 0$ and z_0 is a pole or removable singularity of f. If U is a non-empty open subset of $\mathbb{C}$, then f is meromorphic on U if there is a discrete set S in U such that f is defined and analytic on $U \backslash S$, and all elements of S are poles of f. The meromorphic functions on U form a field.

Definition 4.9. *For f with a Laurent expansion $\sum_{n=-\infty}^{\infty} a_n(z-z_0)^n$ converging on $D^0(z_0, r)$, we define the residue of f at z_0 as*

$$\textit{res}(z_0, f) := a_{-1}.$$

Note that by Theorem 4.7, we have

$$\operatorname{res}(z_0, f) = \frac{1}{2\pi i} \oint_{\gamma_{z_0,r}} f$$

for $0 < r < R$ and in particular, if f is analytic, then $\operatorname{res}(z_0, f) = 0$.

The notion of residues leads to the following remarkable theorem, which is extremely beneficial in computing integrals over closed curves.

Theorem 4.10 (Residue Theorem)**.** *Let γ be a contour in $\mathbb{C}$ and $z_1, \dots, z_p$ be points in the interior of γ. Let f be a complex function that is analytic on an open set containing γ and the interior of γ minus $\{z_1, \dots, z_p\}$. Then the following holds.*

$$\frac{1}{2\pi i} \oint_{\gamma} f = \sum_{i=1}^{p} \textit{res}(z_i, f).$$

This theorem tells us that calculating a contour integral is all about finding poles in the interior and summing up their residues.

In Theorem 5.1, it is suggested that the analytic continuation of the Riemann zeta function to $\mathbb{C}\backslash\{1\}$ has a simple pole at $s = 1$ with residue 1. This can be explained by the following expression for $\zeta(s)$ with Re $s > 1$.

$$\zeta(s) = \frac{\xi(s)\pi^{s/2}\Gamma(\frac{1}{2}s + 1)^{-1}}{s - 1}, \tag{4.1}$$

where $\xi(s) = \frac{1}{2}s(s-1)\pi^{-s/2}\Gamma(\frac{1}{2}s)\zeta(s)$. This function has an analytic continuation to $\mathbb{C}$ and the functional equation $\xi(1-s) = \xi(s)$ for $s \in \mathbb{C}$ and $\xi(0) = \xi(1) = \frac{1}{2}$. For a proof of this, see for example *Multiplicative Number theory* of Davenport [11]. Equation (4.1) then shows that $\zeta(s)$ has an analytic continuation to $\mathbb{C}\backslash\{1\}$. As $s - 1$ is not allowed to be equal to zero in (4.1), we see $s = 1$ is a pole of order 1. Estimating the residue is now done by

$$\begin{aligned}
\lim_{s \to 1}(s-1)\zeta(s) &= \lim_{s \to 1} \xi(s)\pi^{s/2}\Gamma(\tfrac{1}{2}s + 1)^{-1} \\
&= \xi(1)\pi^{1/2}\Gamma(\tfrac{3}{2})^{-1} \\
&= \tfrac{1}{2}\pi^{1/2}(\tfrac{1}{2}\Gamma(\tfrac{1}{2}))^{-1} \\
&= \tfrac{1}{2}\pi^{1/2} \cdot 2\pi^{-1/2} \\
&= 1.
\end{aligned}$$

In the next section, we will elaborate more on the Riemann zeta function and give a formal definition.

Recall the definition of $F_2(t)$ in equation (9.21) from the beginning of this section. The integral can now be calculated using the pole of the Riemann zeta function at $s = 1$ with residue 1. This is explained in Section 9.1.

5 Riemann zeta function

One of the most important functions in analytic number theory is the Riemann-zeta function. As we will use the Riemann zeta function thoroughly in this thesis, an introduction on this function can not be omitted. We first discuss some important properties. Second, we elaborate on the Lindelöf hypothesis for the Riemann-zeta function, which is used in Theorem 1.1 and Theorem 1.2. The proofs and a wider background on the Riemann zeta function can be found in several analytic number theory books, for example in Titchmarsh [43].

5.1 Properties and useful theorems

Let $n \in \mathbb{Z}_{\geq 1}$. We define the Riemann zeta function as

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}, \tag{5.1}$$

where $s = \sigma + it$ is a complex variable. It defines an analytic function on $\{s \in \mathbb{C} : \text{Re } s > 1\}$ and has an analytic continuation to $\mathbb{C}\backslash\{1\}$.

We can also define $\zeta(s)$ as a product

$$\zeta(s) = \prod_{p} (1 - p^{-s})^{-1}, \tag{5.2}$$

where p runs through all the primes. This is called the *Euler product*, from which we can see it connects properties of primes with properties of $\zeta(s)$. One can prove that this series converges for $\sigma > 1$. We also see from this definition, that $\zeta(s)$ has no zeroes for $\sigma > 1$, as a convergent infinite product of non-zero factors is non-zero. The zeros of $\zeta(s)$ lie at $s = -2, -4, -6, \ldots$ and in the *critical strip* $\{s \in \mathbb{C} : 0 < \text{Re } s < 1\}$.

We can see the Riemann zeta function as a Dirichlet L-series, namely as

$$D_1(s) = \sum_{n=1}^{\infty} 1 \cdot n^{-s};$$

the L-function of the principal character modulo 1.

There exist several interesting formulae concerning the Riemann zeta function, which show relations between $\zeta(s)$ and the Möbius function, divisor functions and the Von Mangoldt function. We have seen one relation in (4.1) in Section 4 already. Another very interesting relation between $\zeta(s)$ and the Gamma function is the following.

Theorem 5.1. *[43, p. 13] The function $\zeta(s)$ is analytic for all s except for $s = 1$, where it has a simple pole with residue 1. Furthermore, $\zeta(s)$ satisfies the functional equation*

$$\zeta(s) = 2^s \pi^{s-1} \sin\left(\tfrac{1}{2}s\pi\right) \Gamma(1-s)\zeta(1-s). \tag{5.3}$$

Sometimes we write this as

$$\zeta(s) = \chi(s)\zeta(1-s),$$

where

$$\chi(s) = 2^s \pi^{s-1} \sin(\frac{1}{2}s\pi)\Gamma(1-s).$$

This is often called the symmetry property of $\zeta(s)$. There are several proofs for the functional equation. In Section 4, we computed the residue of the simple pole to be equal to 1. For the functional equation, we give a proof that concerns the Poisson summation formula that is given in Theorem 3.5 and that uses Jacobi's θ-function, which is defined by

$$\theta(\tau) = \sum_{n\in\mathbb{Z}} e^{\pi i n^2 \tau}.$$

The sum is absolutely convergent on Im $\tau > 0$ and thus holomorphic on the upper half plane. It also satisfies two functional equations or symmetries, namely $\theta(\tau+2) = \theta(\tau)$ (periodic modulo 2) and for all $y \in \mathbb{R}_{>0}$ we have

$$\theta(i/y) = \sqrt{y}\theta(iy).$$

The proof of Theorem 5.1 is based on the proof given in [38]. We use some techniques introduced in Section 4.

Proof of Theorem 5.1. Let

$$F(s) = \pi^{-s}\Gamma(s)\zeta(2s),$$

which is meromorphic on $\mathbb{C}$ and holomorphic on Re $s > \frac{1}{2}$ by the properties of $\Gamma(s)$ and $\zeta(s)$. We can write

$$\begin{aligned} F(s) &= \pi^{-s}\Gamma(s)\sum_{n\geq 1} n^{-2s} \\ &= \sum_{n\geq 1} (\pi n^2)^{-s}\Gamma(s) \\ &= \sum_{n\geq 1} \int_0^\infty (\pi n^2)^{-s} t^{s-1} e^{-t} dt \end{aligned}$$

since $\Gamma(s)$ is a Mellin transform. Furthermore, let $t = \pi n^2 y$ and $dt = \pi n^2 dy$. Then $F(s)$ can be expressed as

$$\begin{aligned} F(s) &= \sum_{n\geq 1} \int_0^\infty (\pi n^2)^{-s} (\pi n^2 y)^{s-1} e^{-\pi n^2 y} \pi n^2 dy \\ &= \sum_{n\geq 1} \int_0^\infty y^{s-1} e^{-\pi n^2 y} dy. \end{aligned}$$

Since the sum converges absolutely for Re $s > \frac{1}{2}$, that the sum and integral can be interchanged here. We find

$$F(s) = \int_0^\infty y^{s-1} \sum_{n\geq 1} e^{-\pi n^2 y}.$$

Recall that

$$\theta(iy) = \sum_{n\in\mathbb{Z}} e^{-\pi n^2 y} = 1 + 2\sum_{n\geq 1} e^{-\pi n^2 y}.$$

We obtain

$$\begin{aligned}
F(s) &= \frac{1}{2}\int_0^\infty y^{s-1}(\theta(iy)-1)dy \\
&= \frac{1}{2}\left(\int_0^1 y^{s-1}\theta(iy)dy - \frac{1}{s} + \int_1^\infty y^{s-1}(\theta(iy)-1)dy\right).
\end{aligned}$$

Substituting $t = 1/y$ for the first part of this expression gives $dy = -1/t^2 dt$ and we find that the first part is equal to

$$\begin{aligned}
\int_0^1 y^{s-1}\theta(iy)dy &= \int_\infty^1 t^{1-s}\theta(i/t)\cdot -t^{-2}dt \\
&= \int_1^\infty t^{-s-1}\theta(i/t)dt.
\end{aligned}$$

Furthermore, recall that

$$\theta(i/t) = \sqrt{t}\theta(it).$$

Hence

$$\begin{aligned}
\int_0^1 y^{s-1}\theta(iy)dy &= \int_1^\infty t^{-s-1/2}\theta(it)dt \\
&= \int_1^\infty t^{-s-1/2}(\theta(it)-1)dt + \int_1^\infty t^{-s-1/2}dt \\
&= \int_1^\infty t^{-s-1/2}(\theta(it)-1)dt - \frac{2}{1-2s}.
\end{aligned}$$

We find that $F(s)$ can be written as

$$F(s) = \frac{1}{2}\int_1^\infty (y^{s-1} + y^{-s-1/2})(\theta(iy)-1)dy) - \frac{1}{2s} - \frac{1}{1-2s}.$$

Note that $F(s) = F(\frac{1}{2} - s)$. We can extend F such that it is meromorphic on $\mathbb{C}$. Since Γ has no zeros on Re $s > 0$, the only zeros that F has come from the Riemann zeta function, which has zeros in the critical strip $0 <$ Re $s < 1$. If we look at $F(\frac{s}{2})$, then we see we can also extend $\zeta(s)$ to a meromorphic function on $\mathbb{C}$, since $\Gamma(\frac{s}{2})$ has no zeros on Re $s > 0$ and the simple pole of F at $s = 0$ corresponds to the simple pole of $\Gamma(\frac{s}{2})$ at zero. Since $\Gamma(\frac{s}{2})$ has simple poles at $0, -2, -4, \ldots$, we find that $\zeta(s)$ has simple zeros at $-2, -4, \ldots$, but not at 0. We remove the poles at 0 and 1 by defining

$$\xi(s) = \tfrac{1}{2}s(s-1)F(\tfrac{s}{2}) = \tfrac{1}{2}s(s-1)F(\tfrac{1}{2}(1-s)) = \xi(1-s).$$

Using this expression, we find the functional equation (5.3). $\square$

5.2 Lindelöf Hypothesis for the Riemann zeta function

Bourgain assumes in [5], in this thesis stated in Theorem 1.1, the Lindelöf Hypothesis for the Riemann zeta function. As this concerns the critical strip of the Riemann-zeta function, it is closely related to the Riemann Hypothesis. We first provide some theoretical background on the Lindelöf hypothesis for the Riemann zeta function and finish this section with the hypothesis itself.

For $t \to \infty$, we can find a $k \in \mathbb{N}$ such that $\zeta(\sigma + it) \ll t^k$, see [7, Lemma 2.6.1], for a proof. Hence for $\sigma \in \mathbb{R}$ we can define

$$\mu(\sigma) := \inf\{c : |\zeta(\sigma + it)| \ll t^c\}. \tag{5.4}$$

As a function of σ, it is true that $\mu(\sigma)$ is continuous, non-increasing and convex. By the property that for $\sigma > 1$ we have $|\zeta(s)| \leq \zeta(\text{Re } s)$, we see that $\mu \leq 0$. Also, since $\zeta(s)^{-1} = \sum_{n=1}^{\infty} \frac{\mu(s)}{n^s}$ with $\mu(s)$ the Möbius function, we have that $|\zeta(s)^{-1}| \leq \zeta(\text{Re } s)$, hence

$$\mu(\sigma) = 0 \text{ when } \sigma > 1. \tag{5.5}$$

By the functional equation (5.3), we can also say something about $\mu(\sigma)$ when $\sigma < 0$. Titchmarsh showed in [43, p. 78] that $|\chi(s)| \leq |t/(2\pi)|^{\frac{1}{2}-\sigma}(1 + O(|t|^{-1})$ as $t \to \infty$. Therefore,

$$\mu(s) = \tfrac{1}{2} - \sigma \text{ when } \sigma < 0. \tag{5.6}$$

For $0 < \sigma < 1$, we need the *approximate functional equation* found by Hardy and Littlewood [16], which is

$$\zeta(s) = \sum_{n \leq x} n^{-s} + \chi(s) \sum_{n \leq y} n^{s-1} + O(x^{-\sigma}) + O(|t|^{\frac{1}{2}-\sigma} y^{\sigma-1}) \tag{5.7}$$

for $0 < \sigma < 1$ and $2\pi xy = |t|$. If we let $x = y = (t/(2\pi))^{\frac{1}{2}}$, then

$$\begin{aligned} \zeta(\sigma + it) &\ll x^{1-\sigma} + |t|^{\frac{1}{2}-\sigma} y^{\sigma} \\ &= \left(\frac{t}{2\pi}\right)^{\frac{1}{2}(1-\sigma)} + |t|^{\frac{1}{2}-\sigma} \left(\frac{t}{2\pi}\right)^{\frac{1}{2}\sigma} \\ &\ll |t|^{\frac{1}{2}(1-\sigma)}. \end{aligned}$$

Hence

$$\mu(\sigma) \leq \tfrac{1}{2}(1 - \sigma) \text{ for } 0 \leq \sigma \leq 1. \tag{5.8}$$

Combining (5.5), (5.6) and (5.8) and the fact that $\mu(\sigma)$ is convex, we see that

$$\mu(\sigma) \geq \max(0, \tfrac{1}{2} - \sigma). \tag{5.9}$$

Hypothesis 5.2. *(Lindelöf Hypothesis) We have an equality in* (5.9)*.*

By the convexity of $\mu(\sigma)$ we can rephrase this as $\mu(\frac{1}{2}) = 0$. Several mathematicians have tried to prove the hypothesis, but the smallest bound for $\mu(\frac{1}{2})$ that has been found so far is $\mu(\frac{1}{2}) \leq 13/84$, which is a result by Bourgain [3]. An often used bound is $\mu(\frac{1}{2}) = \frac{1}{6} + \epsilon$, which is proved in [43], Theorem 5.12.

We can also formulate the Lindelöf hypothesis without the notion of $\mu(\sigma)$, namely

$$\zeta(\tfrac{1}{2} + it) = O(t^{\epsilon}) \text{ for every } \epsilon > 0. \tag{5.10}$$

There are several other formulations of the Lindelöf hypothesis. One of them is stated and proved by Backlund, which immediately shows that the Riemann Hypothesis implies the Lindelöf Hypothesis. A proof of this can be found in [43, §13.5].

6 Ergodic theory

The Oppenheim Conjecture [28] was proven by Margulis [24] using results and methods from ergodic theory. Roughly spoken, ergodic theory can be defined as the qualitative study of actions of groups on measure spaces [44], i.e., the long term behaviour. We start this section by giving a short introduction to measure theory, in order to understand the concepts of ergodic theory. Second, we dive into the topic of equidistribution, during which we will also see some ergodic theory.

6.1 Introduction to Measure theory

We discuss some subjects regarding measure theory, such as the Lebesgue measure, Chebyshev's inequality and the Borel-Cantelli Lemma. A good reference for this section is *Real Analysis* of Royden [29]. We start with providing some definitions. Let X be an arbitrary set. The power set of X, denoted by 2^X, is the set of all subsets of X.

Definition 6.1. *Let X and 2^X be as above. A subset $\sigma \subset 2^X$ is called a σ-algebra for X (notation: $\sigma(X)$) if it satisfies the following three properties.*

1. *X is in σ: $X \in \sigma$*
2. *σ is closed under complementation: $A \in \sigma \Rightarrow X \backslash A \in \sigma$.*
3. *σ is closed under countable unions: for $A_1, A_2, \ldots \in \sigma$, where $A_i \cap A_j = \emptyset$ for $i \neq j$, we have $\bigcup_{i=1}^{\infty} A_i \in \sigma$.*

If $A \in \sigma(X)$, then A is called a measurable set in X. We call (X, σ) a measurable space.

Definition 6.2. *Let (X, σ) be a measurable space. On this space, the map $\mu : X \to \bar{\mathbb{R}}$ is called a* measure *if it satisfies the following three properties.*

1. *The null-empty set: $\mu(\emptyset) = 0$.*
2. *Non-negativity: for all $A \in \sigma$ we have $\mu(A) > 0$.*
3. *Countable additivity: for all $A_1, A_2, \ldots \in \sigma$ with $A_i \cap A_j = \emptyset$ for $i \neq j$, we have $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$.*

We call (X, σ, μ) a measure space. Let I be an interval. We define the length $l(I)$, in the usual sense, as the difference of the endpoints of the interval.

Definition 6.3. *Let $A \subset \mathbb{R}$. Let $\{I_n\}$ be the countable collections of open intervals that cover A, i.e., collections for which we have $A \subset \bigcup I_n$. We define the* outer measure *m^*A of A to be the infimum of sums of the lengths of the intervals in the collections. In other words,*

$$m^*A = \inf_{A \subset \bigcup I_n} \sum l(I_n).$$

Note that $A \subset B$ implies $m^*A \leq m^*B$. Furthermore, each set consisting of a single point has outer measure zero. One can prove that the outer measure of an interval is its length.

Definition 6.4. *A set E is called measurable if for each set A, we have*

$$m^*A = m^*(A \cap E) + M^*(A \cap E^c).$$

Furthermore, if $m^*E = 0$, then E is measurable. Also, the following theorem holds.

Theorem 6.5 ([29])**.** *The collection $\mathcal{M}$ of measurable sets is a σ-algebra.*

If E is a measurable set, we define the Lebesgue measure μE to be the outer measure of E. A countable collection of sets $\{A_n\}_{n=1}^{\infty}$ is called ascending if $A_n \subset A_{n+1}$ for each n and descending if $A_{n+1} \subset A_n$ for each n. The Lebesgue measure has the following important property.

Proposition 6.6. *The Lebesgue measure has the following continuity properties.*

(i) If $\{A_n\}_{n=1}^{\infty}$ is an ascending collection of measurable sets, then

$$\mu\left(\bigcup A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

(ii) If $\{B_n\}_{n=1}^{\infty}$ is a descending collection of measurable sets and $\mu(B_1) < \infty$, then

$$\mu\left(\bigcap B_n\right) = \lim_{n\to\infty} \mu(B_n).$$

Let I be a bounded interval and c a real constant. Let χ_I be the characteristic function on the interval I, i.e.,

$$\chi_I(x) = \begin{cases} 1 & \text{if } x \in I \\ 0 & \text{if } x \notin I. \end{cases}$$

We define the Lebesgue integral of $c\chi_I(x)$ over $\mathbb{R}$ as

$$\int_{\mathbb{R}} c\chi_I(x) d\mu(x) = c \cdot \mu(I).$$

The $d\mu(x)$ in the integral denotes that our dummy variable of integration is x and that we are integrating with respect to the Lebesgue measure μ. In general, there are several notations for this.

Theorem 6.7. *(General Chebyshev's Inequality) Let (X, σ, μ) be a measure space and let f be a real-valued measurable function defined on X. Let μ be the Lebesgue measure and let g be a real-valued measurable function that is nonnegative and nondecreasing on the range of f. Then, for any real number $t > 0$ and $0 < p < \infty$ we have*

$$\mu(\{x \in X : f(x) \geq t\}) \leq \frac{1}{g(t)} \int_X g(f(x)) d\mu(x).$$

Proof. Let t be fixed and let $A_t = \{x \in X : f(x) \geq t\}$. Let χ_{A_t} be the characteristic function on A_t. By assumption g is nondecreasing and it is nonnegative on the range of f. Hence

$$0 \leq g(t)\chi_{A_t} \leq g(f(x))\chi_{A_t}.$$

Using Lebesgue integration to integrate over X we get

$$g(t)\mu(A_t) = \int_X g(t)\chi_{A_t} d\mu.$$

By Proposition 2 of [29], the Lebesgue integral is nonnegative and preserves inequalities. Hence

$$\int_X g(t)\chi_{A_t} d\mu \leq \int_X g(f(x))\chi_{A_T} d\mu = \int_{A_t} g(f(x)) d\mu \leq \int_X g(f(x)) d\mu.$$

We find

$$\mu(A_t) \leq \frac{1}{g(t)} \int_X g(f(x)) d\mu.$$

□

Chebyshev's inequality is often stated in a probabilistic form. For completeness, we state this theorem as well. We assume the reader is familiar with the basics of probability theory.

Theorem 6.8. *Let $X : \Omega \to \mathbb{R}$ be a random variable on a probability space $(\Omega, \mathcal{F}, P)$. Suppose X has finite expected value m and finite nonzero variance σ^2. Then, for any real number $k > 0$, we have*

$$P(|X - m| \geq k\sigma) \leq \frac{1}{k^2}.$$

Another often used lemma in measure theory is the Borel-Cantelli Lemma, which uses the properties of Proposition 6.6 in its proof.

For a measurable set A, we say that a property holds for almost everywhere on A, or for almost all $x \in A$, if there is a subset A_0 of A for which the Lebesgue measure $\mu(A_0) = 0$ and the property holds for all $x \in A \backslash A_0$.

Lemma 6.9 (Borel-Cantelli)**.** *Let $\{A_n\}_{n=1}^{\infty}$ be a countable collection of measurable sets for which $\sum_{n=1}^{\infty} A_n < \infty$. Then almost all $x \in \mathbb{R}$ belong to at most finitely many of the A_n's.*

Proof. By the properties of a measurable space, we have for every k,

$$\mu\left(\bigcup_{n=k}^{\infty} A_n\right) \leq \sum_{n=k}^{\infty} \mu(A_n) < \infty.$$

Then by Proposition 6.6, we have

$$\mu\left(\bigcap_{k=1}^{\infty}\left(\bigcup_{n=k}^{\infty} A_n\right)\right) = \lim_{k\to\infty} \mu\left(\bigcup_{n=k}^{\infty} A_n\right) \leq \lim_{k\to\infty} \sum_{n=k}^{\infty} \mu(A_n) = 0.$$

We find that almost all $x \in \mathbb{R}$ fail to belong to $\cap_{k=1}^{\infty}(\cup_{n=k}^{\infty} A_n)$ and therefore belong to at most finitely many A_n's. □

The definition of ergodic uses the notion of measure-preserving transformations. We first define what such transformation is.

Definition 6.10 (Definition 1.1 of [44])**.** *Let* (X_1, σ_1, μ_1) *and* (X_2, σ_2, μ_2) *be measure spaces.*

1. *A transformation* $T : X_1 \to X_2$ *is called measurable if* $T^{-1}(\sigma_2) \subset \sigma_1$.
2. *A transformation* $T : X_1 \to X_2$ *is called measure-preserving if* T *is measurable and if* $\mu_1(T^{-1}(A_2)) = \mu_2(A_2)$ *for all* $A_2 \in \sigma_2$.
3. *A transformation* $T : X_1 \to X_2$ *is called invertible measure-preserving if* T *is measure-preserving, bijective, and* T^{-1} *is also measure preserving.*

Let (X, σ, μ) be a measure space. A transformation $T : X \to X$ is called *ergodic* if it is a measure-preserving transformation, and the only members A_n of σ where $T^{-1}(A_n) = A_n$ satisfy $\mu(A_n) = 0$ or $\mu(A_n) = 1$. We are now ready to state Birkhoff's ergodic theorem. A proof can be found in [44].

Theorem 6.11 (Birkhoff's Ergodic Theorem)**.** *Suppose* T *is a measure-preserving transformation on* (X, σ, μ) *and let* $f \in L^1(\mu)$, *where*

$$L^1(\mu) := \{f : X \to \mathbb{C} | f \text{ measurable and } \int |f| d\mu < \infty\}.$$

Then

$$\frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x))$$

converges almost everywhere to a function $\int f^* d\mu$ *where* $f^* \in L^1(\mu)$. *Also,* $f^* \circ T = f^*$ *almost everywhere, and if* $\mu(X) < \infty$, *then*

$$\int f^* d\mu = \int f d\mu.$$

In the next section, it will become clear how the distribution of a sequence is related to this theorem.

6.2 Uniform distribution mod one

In Section 10, we study the distribution of values of a Diophantine inequality. In this section, we give some background on the distribution of sequences with irrational coefficients.

We study the behaviour of the rational part of a sequence x_n of real numbers in $\mathbb{R}/\mathbb{Z}$, i.e., mod one. We call a sequence uniformly distributed modulo 1, or *equidistributed*, if for every a, b with $0 \leq a < b < 1$ we have

$$\frac{1}{n}\#\{j : 0 \leq j \leq n-1, \{x_j\} \in [a, b]\} \to b - a \text{ as } n \to \infty.$$

Here $\{x_n\}$ is the usual notation for the fractional part of a real number. The following result is a well known condition for x_n to be uniformly distributed mod one.

Theorem 6.12 (Weyl's criterion [45])**.** *A sequence x_n of real numbers is uniformly distributed modulo one if and only if for every integer $b \neq 0$ we have*

$$\frac{1}{n}\sum_{j=0}^{n-1} e(bx_j) \to 0$$

as $n \to \infty$.

Note that this expression is exactly the problem of non-trivially estimating an exponential sum; an important problem in analytic number theory. We deduce the following.

Corollary 6.13. *A sequence x_n of real numbers is uniformly distributed mod one if and only if for any continuous function $f : [0,1] \to \mathbb{R}$ with $f(0) = f(1)$ we have*

$$\frac{1}{n}\sum_{j=0}^{n-1} f(\{x_j\}) \to \int_0^1 f(x)dx.$$

Proof. Let x_j be uniformly distributed mod one. Let $\chi_{[a,b]}$ be the characteristic function of the interval $[a,b]$. Then the definition of uniform distribution can be rewritten as

$$\frac{1}{n}\sum_{j=0}^{n-1} \chi_{[a.b]}(x_j) \to \int_0^1 \chi[a,b](x)dx$$

as $n \to \infty$. This implies that for g being a step function, i.e., it is a finite linear combination of characteristic functions of intervals, that

$$\frac{1}{n}\sum_{j=0}^{n-1} g(x_j) \to \int_0^1 g(x)dx$$

as $n \to \infty$. Let f be a continuous function on $[0,1]$. We can find a step function, given $\epsilon > 0$, such that $||f - g|| \leq \epsilon$. Hence

$$\begin{aligned}\left|\frac{1}{n}\sum_{j=0}^{n-1} f(x_j) - \int_0^1 f(x)dx\right| &\leq \left|\frac{1}{n}\sum_{j=0}^{n-1}(f(x_j) - g(x_j))\right| + \left|\frac{1}{n}\sum_{j=0}^{n-1} g(x_j) - \int_0^1 g(x)dx\right| \\ &+ \left|\int_0^1 g(x)dx - \int_0^1 f(x)dx\right| \\ &\leq \left|\frac{1}{n}\sum_{j=0}^{n-1} g(x_j) - \int_0^1 g(x)dx\right| + 2\epsilon.\end{aligned}$$

As the last term converges to 0 for $n \to \infty$, we find

$$\limsup_{n\to\infty} \left|\frac{1}{n}\sum_{j=0}^{n-1} f(x_j) - \int_0^1 f(x)dx\right| \leq 2\epsilon,$$

hence

$$\frac{1}{n}\sum_{j=0}^{n-1} f(x_j) \to \int_0^1 f(x)dx$$

as $n \to \infty$. □

Sometimes, the little-oh notation is being used. The statement in Theorem 6.12 will then be replaced with $\sum_{j=1}^{n-1} e(bx+j) = o_b(n)$ as $n \to \infty$ [15].

Example 6.14. The sequence $\{\log(n)\}_{n>1}$ is not uniformly distributed modulo one. Let $b = 1$. Note

$$\frac{1}{n}\sum_{j=1}^{n} e^{2\pi i \log(j)} = \frac{1}{n}\sum_{j=1}^{n} j^{2\pi i}.$$

We use Euler's summation formula, which states that for $f(t)$ a complex valued function with continuous derivative on $t \in \{1, N\}$, with $N \geq 1$ an integer, we have

$$\sum_{n=1}^{N} f(n) = \int_1^N f(t)dt + \frac{1}{2}(f(1)+f(N)) + \int_1^N (\{t\} - \tfrac{1}{2})f'(t)dt.$$

Let $f(t) = e^{2\pi i \log t}$. Divide both sides of the equation by N. Then

$$\frac{1}{N}\int_1^N f(t)dt = \frac{1}{N}\int_1^N t^{2\pi i}dt = \frac{1}{2\pi i N} t^{2\pi i + 1} \left.\right|_{t=1}^{N},$$

which does not converge for $N \to \infty$,

$$\frac{1}{N}\frac{1}{2}\left(f(1) - f(N)\right) = \frac{1}{2N} + \frac{1}{2}N^{2\pi i - 1} \to 0$$

as $N \to \infty$, and

$$\left| \int_1^N (\{t\} - \tfrac{1}{2}) \cdot 2\pi i t^{2\pi i - 1} dt \right| \leq \frac{1}{2} \cdot 2\pi \int_1^N \frac{1}{t}dt.$$

Hence the third term converges to zero as well, for $N \to \infty$. We obtain

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(n) \to \infty.$$

Weyl's criterion is violated and therefore, $\{\log n\}_{n\geq 1}$ is not uniformly distributed modulo one.

The next famous corollary is proved by Weyl's criterion as well.

Corollary 6.15. *The sequence $x_n = \alpha n$ is uniformly distributed modulo one if and only if $\alpha \in \mathbb{R}\backslash\mathbb{Q}$.*

Proof. Let $\alpha \in \mathbb{Q}$, say $\alpha = p/q$, where $p \in \mathbb{Z}, q \in \mathbb{N}$, $\gcd(p,q) = 1$. Then $\{\alpha\}$ can only take q distinct values, namely

$$0, \left\{\frac{p}{q}\right\}, \left\{\frac{2p}{q}\right\}, \ldots, \left\{\frac{(q-1)p}{q}\right\}.$$

As this number is finite, we see it cannot be distributed uniformly modulo one. Assume now $\alpha \in \mathbb{R}\backslash\mathbb{Q}$. As $e(b\alpha) \neq 1$ for b a nonzero integer, we have

$$\frac{1}{n}\sum_{j=0}^{n-1} e(b\alpha j) = \frac{1}{n}\frac{e(b\alpha n) - 1}{e(b\alpha) - 1}.$$

This implies

$$\left| \frac{1}{n} \sum_{j=0}^{n-1} e(b\alpha j) \right| \leq \frac{1}{n} \frac{2}{|e(b\alpha) - 1|} \to 0$$

as $n \to \infty$. By Theorem 6.12, αn is uniformly distributed modulo one. □

As we can linearly transform αn to a sequence $y_n = \alpha n + \beta$, we see y_n is uniformly distributed modulo one if and only if α is irrational as well. Corollary 6.15 is often called *Weyl's equidistribution theorem.*

Remark 6.16. Weyl's equidistribution theorem implies Kronecker's theorem, which says that for α an irrational number, the sequence $\{n\alpha\}_{n\geq 1}$ is dense in $[0, 1)$.

Remark 6.17. Birkhoff's Ergodic Theorem 6.11 implies Weyl's equidistribution theorem. In fact, Weyl's theorem can be seen as a special case of Birkhoff's theorem.

One might ask what can be said about higher degree polynomials in n. We will prove the following theorem, which is also a theorem by Weyl.

Theorem 6.18. *Let α be an irrational real number. Then the sequence*

$$\{\alpha n^2 : n \geq 1\}$$

is uniformly distributed modulo one.

In order to prove this, we need the following lemma.

Lemma 6.19. *Let $a < b$ be nonnegative integers. Let θ be an irrational real number. Then*

$$\left| \sum_{n=a}^{b} e(n\theta) \right| \ll \min(b - a, \frac{1}{||\theta||}).$$

Proof. Recall that $|\sin(t)| \leq ||t||$. This implies $|e(t) - 1| \leq \pi||t||$. Using the triangle inequality and the fact that a, b are integers, we find

$$\left| \sum_{n=a}^{b} e(n\theta) \right| \leq a - b + 1 \ll a - b.$$

As the exponential function is a geometric series, we have

$$\begin{aligned} \left| \sum_{n=a}^{b} e(n\theta) \right| &= \frac{|e(a\theta) - e((b+1)\theta)|}{|1 - e(\theta)|} \\ &\leq \frac{2}{|1 - e(\theta)|} \\ &= \frac{2}{|e(\frac{\theta}{2}) - e(\frac{-\theta}{2})|} = \frac{1}{|\sin(\pi\theta)|}. \end{aligned}$$

As $|\sin(\pi t)| \geq 2||t||$ for all t, we obtain

$$\left| \sum_{n=a}^{b} e(n\theta) \right| \leq \frac{1}{2||\theta||}.$$

This completes the proof. □

In the proof of Theorem 6.18, we also use the well-known theorem of Dirichlet, which states that for $\alpha \in \mathbb{R}\backslash\mathbb{Q}$, there exists infinitely many pairs of integers p, q, with $\gcd(p, q) = 1$ and $2 \leq q$ such that $|\alpha - \frac{p}{q}| \leq \frac{1}{q^2}$.

Proof of Theorem 6.18. In this proof, we make use of *Weyl differencing.* Define $S = \sum_{n=0}^{N} e(n^2\alpha)$. Then

$$|S|^2 = \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} e(\alpha(n_1^2 - n_2^2)).$$

The idea of Weyl differencing is that this squared sum can be written as a polynomial of one degree lower, i.e., the exponent does not contain $(n_1 - n_2)^2$ anymore, but a linear polynomial in n_1, n_2. This can be done by re-indexing this sum. We set $h = n_1 - n_2$ such that $-N \leq h \leq N$ and $\max(0, -h) \leq n_2 = n_1 - h \leq \min(N, N - h)$. Then

$$|S|^2 = \sum_{h=-N}^{N} \sum_{n_2=\max(-h,0)}^{\min(N,N-h)} e(\alpha(2hn_2+h^2)) = \sum_{h=-N}^{N} e(\alpha h^2) \sum_{n_2=\max(-h,0)}^{\min(N,N-h)} e(\alpha(2hn_2)).$$

Using triangle inequality and Lemma 6.19 we obtain

$$|S|^2 \ll \sum_{h=-N}^{N} \min(N, \frac{1}{||2h\alpha||}).$$

Divide the interval $[-N, N]$ into smaller intervals of length at most $\frac{q}{2}$, each of the form $M \leq h < M + \frac{q}{2}$.

Claim: The sum of $\min(N, \frac{1}{||2h\alpha||})$ over each interval is $\ll N + q \log q$.

We will prove the claim for $M = 0$; the proof for other values of M is similar. Define $\tilde{S} = \sum_{0 \leq h < q/2} \min(N, \frac{1}{||2h\alpha||})$. Also, by Dirichlet, write $\alpha = p/q + \theta$ where $|\theta| \leq 1/q^2$, $\gcd(p, q) = 1$. As $0 \leq 2h < q$, all residues of $2h$ mod q are distinct, so $2hp$ is congruent to $0, 1$ and -1 mod q at most once. For other values of h, we have $||2h\alpha|| \geq \left\| \frac{2hp}{q} \right\| - \frac{2h}{q^2} > 0$. Hence

$$\tilde{S} \leq 3N + \sum_{\substack{0 \leq h < q/2 \\ 2hp \not\equiv 0,1,-1 \bmod q}} \min\left(N, \frac{1}{\left\|\frac{2hp}{q}\right\| - \frac{1}{q}}\right).$$

As $\|2hpq^{-1}\|$ takes on each value of $\frac{2}{q}, \ldots, \frac{\lfloor q/2 \rfloor}{q}$ at most twice, we obtain

$$\tilde{S} \leq 3N + 2\sum_{j=2}^{\lfloor q/2 \rfloor} \frac{1}{j/q - 1/q} = 3N + 2q \sum_{j=1}^{\lfloor q/2 \rfloor - 1} j^{-1} \ll N + q \log q.$$

This proves the claim. As there are $\ll N/q + 1$ intervals, we find

$$|S|^2 \ll (N + q \log q)(N/q + 1) \ll N^2/q + (q + N) \log q.$$

Then

$$\frac{1}{N} \left| \sum_{n=1}^{N} e(n^2\alpha) \right| \ll \frac{1}{\sqrt{q}} + \sqrt{\frac{q \log q}{N^2} + \frac{\log q}{N}}$$

and this converges to $1/\sqrt{q}$ as $N \to \infty$. By Dirichlet's theorem, q can be arbitrarily large. So if we take $q \to \infty$, this sum converges to zero and thus αn^2 is uniformly distributed mod one by Theorem 6.12. $\square$

We can generalise this theorem into the following.

Theorem 6.20. *Let $P(x) = a_d x^d + a_{d-1}x^{d-1} + \ldots + a_1 x + a_0$ be a polynomial with at least one of the coefficients $a_1, \ldots, a_d$ irrational. Then the sequence $\{P(n) : n \geq 1\}$ is uniformly distributed modulo one.*

Another generalisation of Weyl's criterion is about applying the criterion on k dimensions. Given $v = (a_1, \ldots, a_k) \in \mathbb{R}^k$, we define $v \mod 1$ to be the vector $(a_1 \mod 1, \ldots, a_k \mod 1)$. A sequence of vectors $v_1, v_2, \ldots \in \mathbb{R}^k$ is called uniformly distributed mod one if for any $0 \leq b_j < c_j < 1$ for $j = 1, 2, \ldots, k$ we have

$$\# \left\{ n \leq N : a_n \mod 1 \in \bigoplus_{j=1}^{k} [b_j, c_j) \right\} \sim \prod_{j=1}^{k} (c_j - b_j) \cdot N$$

as $N \to \infty$ [15].

Theorem 6.21. *A sequence of vectors $v_1, v_2, \ldots, \in \mathbb{R}^k$ is uniformly distributed mod one if and only if for every $b \in \mathbb{Z}^k$, $b \neq 0$ we have*

$$\frac{1}{n} \sum_{j=0}^{n-1} e(bv_j) \to 0$$

as $n \to \infty$.

A famous corollary from Kronecker [20], is that if $1, \alpha_1, \ldots, \alpha_k$ are linearly independent over $\mathbb{Q}$, then $\{(n\alpha_1, n\alpha_2, \ldots, n\alpha_k) : n \geq 1\}$ are uniformly distributed mod one. Furthermore, from Theorem 6.20 and 6.21, one can deduce that the vectors $\{(n\alpha, n^2\alpha, \ldots, n^k\alpha) : n \geq 1\}$ are uniformly distributed mod one if $\alpha \in \mathbb{R} \backslash \mathbb{Q}$ [15]. More on equidistribution of polynomial sequences in $\mathbb{R}/\mathbb{Z}$ can be found in *Higher Order Fourier Analysis* from T. Tao [41].

These theorems look very promising for proving the forms used in Section 9 are uniformly distributed. However, a sequence that is uniformly distributed modulo one does not have to be uniformly distributed in general. Furthermore, these theorems state behaviour for $n \to \infty$, but do not tell us *when* this will happen exactly. In other words, Weyl's condition and its corollaries are qualitative, not quantitative. A quantitative version of Weyl's criterion is the Erdös-Turán-Koksma inequality [12].

A lot of research has been done on quantitative results on distributions of, for example, quadratic forms. For example, the distribution of $\alpha m^2 + n^2$, α irrational, is explored in [2]. In this article, it is assumed that the sequence is randomly distributed. The goal is to bound the smallest gap between two points in the sequence. It turns out that the order of growth is consistent with Poisson statistics, but finer details of the Poisson distribution are violated.

Recall that a Poisson probability distribution is applied when we look at the number of times an event occurs in a certain time interval, where events occur randomly and independently [40]. Define λ as the mean number of occurrences of the event in the given interval. Then

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}.$$

Its mean is equal to λ. If N points are picked independently and uniformly in $[0, N]$, it is believed that the smallest gap is almost surely of size $\approx 1/N$ [22]. In Section 10, we look at the proof of the smallest gap for the sequence $\alpha m^2 + n^2$ with $\alpha > 0$ irrational, based on [2].

We finish this section with the notion that other articles with quantitative research on distributions, such as [2], [32] and [30], also contribute to understanding the distribution of quadratic forms and forms in higher dimensions. Popular statistical measures are the distribution of gaps between consecutive solutions and the variance of solutions, which use a so-called pair correlation function.

7 Implicit Function Theorem

Solutions of a quadratic equation $ax^2 + bx + c = 0$, can be expressed as $x = \frac{1}{2a}(-b \pm \sqrt{b^2 - 4ac})$. For non-quadratic equations, with more variables, it is sometimes not straightforward how to express a variable in terms of the other variables. In this section, we explore how we can proceed for the more difficult equations. We introduce the Implicit Function Theorem, which will be used in Section 9.

Suppose we have a system of nonlinear equations $F_i : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $F_i(\mathbf{x}, \mathbf{y}) = F_i(x_1, \ldots, x_n, y_1, \ldots, y_m)$, $1 \leq i \leq m$. In most cases, we are not able to solve this system uniquely. However, we could try to find a relation between y_i and $F_i(\mathbf{x}, \mathbf{y})$, i.e., we write y_i as a function of $(x_1, \ldots, x_n)$. Assume therefore that we have a solution

$$F_i(\mathbf{x}^0, \mathbf{y}^0) = F_i(x_1^0, \ldots, x_n^0, y_1^0, \ldots, y_m^0) = 0,\ 1 \leq i \leq m.$$

The question is: when can we find, for each $\mathbf{x} = (x_1, \ldots, x_n)$ near this solution $\mathbf{x}^0$ a unique $\mathbf{y} = (y_1, \ldots, y_m)$ near $\mathbf{y}^0$, which satisfies $F_i(\mathbf{x}, \mathbf{y}) = 0$? The answer is given by the Implicit Function Theorem. There are several ways of stating the theorem. Here, we state the theorem given by M. Spivak in [36].

Theorem 7.1 (Implicit Function Theorem)**.** *Suppose $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is continuously differentiable in an open set containing $(\mathbf{x}^0, \mathbf{y}^0)$ and suppose $F(\mathbf{x}^0, \mathbf{y}^0) = 0$. Let M be the $m \times m$ matrix given by*

$$(D_{n+j}F_i(\mathbf{x}^0, \mathbf{y}^0)),\ 1 \leq i, j \leq m.$$

In other words,

$$M = \begin{pmatrix} \frac{\partial f^1}{\partial y_1^0} & \cdots & \frac{\partial f^1}{\partial y_m^0} \\ \vdots & \ddots & \vdots \\ \frac{\partial f^m}{\partial y_1^0} & \cdots & \frac{\partial f^m}{\partial y_m^0} \end{pmatrix}.$$

If $\det M \neq 0$, then there is an open set $A \subset \mathbb{R}^n$ containing $\mathbf{x}^0$ and an open set $B \in \mathbb{R}^m$ containing $\mathbf{y}^0$, such that for each $\mathbf{x} \in A$ there is a unique $g(\mathbf{x}) \in B$ such that $F(\mathbf{x}, g(\mathbf{x})) = 0$. Furthermore, the function g is differentiable.

Note that if $F(\mathbf{x}, \mathbf{y}) = c$, we can always define $G(\mathbf{x}, \mathbf{y}) = F(\mathbf{x}, \mathbf{y}) - c$ and use the Implicit Function Theorem for $G(\mathbf{x}, \mathbf{y})$. This will be done in Section 9. The Implicit Function Theorem is strongly related to another well-known theorem, the Inverse Function Theorem. We will state this theorem below. A proof can be found in [36].

Theorem 7.2 (Inverse Function Theorem)**.** *Suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable in an open set containing $\mathbf{x}^0$, and $\det f'(\mathbf{x}^0) \neq 0$. Then there is an open set V containing $\mathbf{x}^0$ and an open set W containing $f(\mathbf{x^0})$ such that $f : V \to W$ has a continuous inverse $f^{-1} : W \to V$ which is differentiable and for all $\mathbf{y} \in W$ we have*

$$(f^{-1})'(\mathbf{y}) = [f'(f^{-1}(\mathbf{y}))]^{-1}.$$

We will use this theorem to prove Theorem 7.1.

Proof of Theorem 7.1. We define $\Omega : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$ by $\Omega(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, F(\mathbf{x}, \mathbf{y}))$. Then $\det \Omega'(\mathbf{x}^0, \mathbf{y}^0)$ is given by

$$\det \begin{pmatrix} \frac{\partial x_1^0}{\partial x_1^0} & \cdots & \frac{\partial x_1^0}{\partial x_n^0} & \frac{\partial x_1^0}{\partial y_1^0} & \cdots & \frac{\partial x_1^0}{\partial y_m^0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n^0}{\partial x_1^0} & \cdots & \frac{\partial x_n^0}{\partial x_n^0} & \frac{\partial x_n^0}{\partial y_1^0} & \cdots & \frac{\partial x_n^0}{\partial y_m^0} \\ \frac{\partial y_1^0}{\partial x_1^0} & \cdots & \frac{\partial y_1^0}{\partial x_n^0} & \frac{\partial y_1^0}{\partial y_1^0} & \cdots & \frac{\partial y_1^0}{\partial y_m^0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m^0}{\partial x_1^0} & \cdots & \frac{\partial y_m^0}{\partial x_n^0} & \frac{\partial y_m^0}{\partial y_1^0} & \cdots & \frac{\partial y_m^0}{\partial y_m^0} \end{pmatrix} = \det \begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \frac{\partial f^1}{\partial y_1^0} & \cdots & \frac{\partial f^1}{\partial y_m^0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{\partial f^m}{\partial y_1^0} & \cdots & \frac{\partial f^m}{\partial y_m^0} \end{pmatrix}.$$

This is a triangular block matrix, which means $\det \Omega'(\mathbf{x}^0, \mathbf{y}^0) = \det M \neq 0$. We can use Theorem 7.2 now, which implies there is an open set $W \subset \mathbb{R}^n \times \mathbb{R}^m$ containing $F(\mathbf{x}^0, \mathbf{y}^0) = (\mathbf{x}^0, \mathbf{0})$, and an open set in $\mathbb{R}^n \times \mathbb{R}^m$ containing $(\mathbf{x}^0, \mathbf{y}^0)$. We take this open set to be of the form $A \times B$, such that $\Omega : A \times B \to W$ has a differentiable inverse $h : W \to A \times B$. Since Ω is of the form $(\mathbf{x}, F(\mathbf{x}, \mathbf{y}))$, we find that h is of the form $h(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, k(\mathbf{x}, \mathbf{y}))$ for some differentiable function k. Let $\pi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be defined by $\pi(\mathbf{x}, \mathbf{y}) = \mathbf{y}$. We have $\pi \circ \Omega = F$. Hence

$$\begin{aligned} F(\mathbf{x}, k(\mathbf{x}, \mathbf{y})) &= F \circ h(\mathbf{x}, \mathbf{y}) = (\pi \circ \Omega) \circ h(\mathbf{x}, \mathbf{y}) \\ &= \pi \circ (\Omega \circ h)(\mathbf{x}, \mathbf{y}) = \pi(\mathbf{x}, \mathbf{y}) = \mathbf{y}. \end{aligned}$$

This means $F(\mathbf{x}, k(\mathbf{x}, \mathbf{0})) = 0$. In other words, we can define a function $g(\mathbf{x}) = k(\mathbf{x}, \mathbf{0})$. □

Since g is differentiable, we can find its derivative. We have $F^i(\mathbf{x}, g(\mathbf{x})) = 0$, which means taking the derivative gives us

$$0 = D_j F^i(\mathbf{x}, g(\mathbf{x})) + \sum_{\alpha=1}^{m} D_{n+\alpha} f^i(\mathbf{x}, g(\mathbf{x})) \cdot D_j g^\alpha(\mathbf{x})$$

for $i, j = 1, \dots, m$. Since $\det M \neq 0$, we can solve these equations for $D_j g^\alpha(\mathbf{x})$. We see that the answer will depend on the various $D_j F^i(\mathbf{x}, g(\mathbf{x}))$ and thus on $g(\mathbf{x})$.

Example 7.3. We use the previous notion in an example. Consider the system

$$\begin{cases} F_1 = x^3 y - z = -1 \\ F_2 = x + y^2 + z^3 = 11 \end{cases}$$

Then $(x, y, z) = (1, 1, 2)$ is a solution of the system. Let us compute the values for x and y when $z = 2.1$. We first compute the matrix M as in Theorem 7.1.

$$\begin{pmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} & \frac{\partial F_1}{\partial z} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} & \frac{\partial F_2}{\partial z} \end{pmatrix} = \begin{pmatrix} 3x^2 y & x^3 & -1 \\ 1 & 2y & 3z^2 \end{pmatrix}.$$

At the solution $(1, 1, 2)$, the matrix is

$$\begin{pmatrix} 3 & 1 & -1 \\ 1 & 2 & 12 \end{pmatrix}.$$

The determinant of the partial derivatives to x and y is

$$\begin{vmatrix} 3 & 1 \\ 1 & 2 \end{vmatrix} = 6 - 1 = 5 \neq 0.$$

We conclude: there exists a solution near $(1, 1, 2)$. Denote by $x(z), y(z)$ the values of x resp. y when the value for z is given. Denote $x'(z)$, $y'(z)$ for their derivatives. To find the derivatives, we solve the equation

$$\begin{pmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y}. \end{pmatrix} \begin{pmatrix} x'(2) \\ y'(2) \end{pmatrix} = \begin{pmatrix} -\frac{\partial F_1}{\partial z} \\ -\frac{\partial F_2}{\partial z} \end{pmatrix} = \begin{pmatrix} 1 \\ -12 \end{pmatrix}.$$

We obtain

$$\begin{pmatrix} 3x'(2) + y'(2) \\ x'(2) + 2y'(2) \end{pmatrix} = \begin{pmatrix} 1 \\ -12 \end{pmatrix}.$$

Hence, $x'(2) = 2\frac{4}{5}, y'(2) = -7\frac{2}{5}$. We conclude

$$\begin{aligned} x(2.1) &\approx x(2) + x'(2) \cdot 0.1 = 1.28 \\ y(2.1) &\approx y(2) + y'(2) \cdot 0.1 = 0.26. \end{aligned}$$

We have found an estimation of another solution $(1.28, 0.26, 2.1)$.

The Implicit Function Theorem is very helpful in multivariable calculus. In the previous example, it is shown how the theorem can be used. However, it is not constructive in general: it only tells us there exists a solution, not how this solution is derived. This will be a stumbling block in Section 9.

8 Dimension Growth Conjecture

We can reformulate the idea of finding integer solutions of $|Q(x)-\xi| < \delta$, taking $\xi = 0$, as introduced in Theorem 1.1 to finding rational points near the planar curve $\mathcal{C} \subset \mathbb{R}^2$ given by $1 + \alpha_2 y_2^2 - \alpha_3 y_3^2$. We will use some ideas and results of a recent paper of Huang [18] in which he gives a sharp asymptotic formula for the number of rational points up to a given height near a hypersurface.

8.1 Prerequisites

This section thoroughly uses concepts regarding algebraic varieties. We state the definitions that we will use in this section. These definitions are taken from Silverman [35]. Let K be a perfect field, which means every algebraic extension of K is separable. We denote the algebraic closure of K by $\bar{K}$.

Definition 8.1. *The affine n-space over K is the set of n-tuples*

$$\mathbb{A}^n(\bar{K}) = \{P = (x_1, \dots, x_n) : x_i \in \bar{K}\}.$$

The set of K-rational points of $\mathbb{A}^n$ is the set

$$\mathbb{A}^n(K) = \{P = (x_1, \dots, x_n) \in \mathbb{A}^n : x_i \in K\}.$$

Let $\bar{K}[X_1, \dots, X_n] =: \bar{K}[X]$ be a polynomial ring in n variables. Let $I \subset \bar{K}[X]$ be an ideal. A subset of $\mathbb{A}^n$ associated to I, defined by

$$V_I := \{P \in \mathbb{A}^n : f(P) = 0 \text{ for all } f \in I\}$$

is called an affine algebraic set.

Definition 8.2. *Let V be an algebraic set. The ideal of V is given by*

$$I(V) := \{f \in \bar{K}[X] : f(P) = 0 \textit{ for all } P \in V\}.$$

We call an algebraic set defined over K if its ideal $I(V)$ can be generated by polynomials in $K[X]$.

Definition 8.3. *An affine algebraic set V is called an affine variety if $I(V)$ is a prime ideal in $\bar{K}[X]$.*

If V is a variety, then the dimension of V, denoted $\dim(V)$, is the transcendence degree of $\bar{K}(V)$ over $\bar{K}$.

We can now give the definition of a projective space.

Definition 8.4. *The projective n-space over K, denoted $\mathbb{P}^n$, is the set of all $(n+1)$-tuples*

$$(x_0, \dots, x_n) \in \mathbb{A}^{n+1}$$

such that at least one x_i is nonzero, modulo the equivalence relation

$$(x_0, \dots, x_n) \sim (y_0, \dots, y_n)$$

if there exists a $\lambda \in \bar{K}^*$, such that $x_i = \lambda y_i$ for all i. Furthermore denote by $[x_0, \dots, x_n]$ the equivalence class $\{\lambda x_0, \dots, \lambda x_n) : \lambda \in \bar{K}^*\}$. We call the individual coordinates $x_0, \dots, x_n$ homogeneous coordinates for the corresponding point in $\mathbb{P}^n$.

We can use a similar definition for a projective algebraic set as we defined for the affine case.

Definition 8.5. *A polynomial* $f \in \bar{K}[X] = \bar{K}[X_0, \dots, X_n]$ *is homogeneous of degree* d *if*

$$f(\lambda X_0, \dots, \lambda X_n) = \lambda^d f(X_0, \dots, X_n)$$

for all $\lambda \in \bar{K}$. *We call an ideal* $I \in \bar{K}[X]$ *homogeneous if it is generated by homogeneous polynomials.*

Finally, a projective algebraic set is called a projective variety if its homogeneous ideal $I(V)$ is a prime ideal in $\bar{K}[X]$. In this section, we assume the varieties to be cut out by a finite system of homogeneous equations over $\mathbb{Q}$ and when we call a variety irreducible, we will mean that the variety is geometrically reduced and irreducible.

8.2 A basic counting function

We focus on counting rational points lying on a manifold itself. Let $f \in \mathbb{Z}[x_1, \dots, x_n]$ be a polynomial and

$$S_f := \{\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{Z}^n \backslash \{0\} : f(\mathbf{x}) = 0\}$$

the corresponding zero locus of non-zero integer solutions. We want to understand how the counting function

$$N(f; B) := \#\{\mathbf{x} \in S_f : ||\mathbf{x}|| \leq B\} \tag{8.1}$$

behaves when $B \to \infty$. The norm $|| \cdot || : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is an arbitrary norm here. Suppose f is a polynomial of degree $d \geq 1$, then for the vectors that are counted in (8.1), the values $f(\mathbf{x})$ are of order B^d and some have exact order B^d. Therefore, we would expect that the probability of a random chosen value of $f(\boldsymbol{x})$ that vanishes is of order B^{-d}. Since we have a polynomial in n variables, we expect that

$$N(f; B) \asymp B^{n-d}.$$

Unfortunately, this is not always the case, as local conditions sometimes imply that $N(f; B)$ is identically zero. Some examples where this might happen can be found in the first chapter of [6]. Therefore, we need some conditions on f for which the heuristic is true. The following theorem, due to Birch [1], gives an asymptotic formula for our counting function which supports the heuristic.

Theorem 8.6. *Suppose* $f \in \mathbb{Z}[x_1, \dots, x_n]$ *is a non-singular homogeneous polynomial of degree* d *in* $n > (d-1)2^d$ *variables. Assume that* $f(\mathbf{x}) = 0$ *has non-trivial solutions in* $\mathbb{R}$ *and each* p*-adic field* $\mathbb{Q}_p$. *Then there is a constant* $c_f > 0$ *such that*

$$N(f; B) \sim c_f B^{n-d}$$

as $B \to \infty$.

The proof uses the Hardy-Littlewood circle method. See Section 11 for a introduction on this method.

We can expand this idea by phrasing this counting function in terms of arbitrary projective algebraic varieties $V \subset \mathbb{P}^{n-1}$. Let $x = [\mathbf{x}] \in \mathbb{P}^{n-1}(\mathbb{Q})$ be a projective rational point with $x \in \mathbb{Z}^n$ chosen in a way that $\gcd(x_1, \ldots, x_n) = 1$. We define the height of x to be $H : \mathbb{P}^{n-1}(\mathbb{Q}) \to \mathbb{R}_{>0}$, $x \mapsto ||\mathbf{x}||$. Given a locally closed subset $U \subset V$, we define

$$N_U(B) := \#\{x \in U(\mathbb{Q}) : H(x) \leq B\}$$

for each $B \geq 1$, to be a more generalised counting function. All known examples of this counting functions are of the shape

$$N_U(B) \sim cB^a(\log B)^b,$$

as $B \to \infty$. Here $a, b, c \geq 0$, $a \in \mathbb{Q}$ and $b \in \frac{1}{2}\mathbb{Z}$. The main difference between $N_U(B)$ and (8.1) is that we now only look at primitive integer solutions. The relation between the two counting functions is given in the equation

$$N_V(B) = \frac{1}{2}\sum_{k=1}^{\infty} \mu(k) N(f; k^{-1}B),$$

where $\mu(k)$ is the Möbius function. This is encountered by the fact that $\mathbf{x}$ and $-\mathbf{x}$ represent the same point in $\mathbb{P}^{n-1}$ and that we have the following identity for the Möbius function:

$$\sum_{d|n} \mu(d) = \begin{cases} 1, & n = 1 \\ 0, & n \in \mathbb{Z}_{>1} \end{cases}$$

8.3 Upper bounds for $N_V(B)$

Let $V \subset \mathbb{P}^{n-1}$ be an irreducible projective variety of degree d again, whose ideal is generated by homogeneous polynomials $f(\mathbf{x})$ defined over the rationals. Then the following holds.

Lemma 8.7. *When V is a linear space, so when $d = 1$, we have the following asymptotic formula.*

$$N_V(B) \sim c_V B^{\dim V + 1}, \tag{8.2}$$

for $B \to \infty$ and for c_V an appropriate constant $c_V > 0$.

Proof. When $V \subset \mathbb{P}^{n-1}$ is a linear space, we can write V as a system of linear equations

$$\begin{cases} a_{11}x_1 + \ldots + a_{1n}x_n = 0 \\ \qquad \cdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n = 0 \end{cases} \tag{8.3}$$

Without loss of generality, we can even assume $V = \mathbb{P}^m$, so $\dim V = m$. Then

$$N_V(B) = \frac{1}{2}\{\mathbf{x} \in \mathbb{Z}^{m+1} \backslash \{0\},\ \ \gcd(x_0, \ldots, x_m) = 1,\ \ \max|x_i| \leq B\}.$$

We see that this number $N_V(B)$ is bounded above by a constant times $B^{m+1} = B^{\dim V+1}$. As we let $B \to \infty$, the number of solutions will grow asymptotically with this bound, hence

$$N_V(B) \sim c_V B^{\dim V + 1}.$$

□

In fact, $N_V(B)$ can never grow faster for arbitrary d than it does for linear varieties. This is stated in the following theorem.

Theorem 8.8. *Let $V \subset \mathbb{P}^{n-1}$ be a variety of degree d. Then*

$$N_V(B) \ll_{d,n} B^{\dim V+1}.$$

A proof of this theorem can be found in [6]. We would like to know if we can improve upon this upper bound. Since this bound is optimal if V contains a linear component of maximal dimension defined over $\mathbb{Q}$, we need to make some assumptions. Furthermore, if V contains a linear divisor which is defined over the rationals, then it contains a space $W \subset V$ such that $\dim W = \dim V - 1$. Hence

$$N_V(B) \gg_V B^{\dim V}.$$

In 2013, Serre [34] stated the following conjecture.

Conjecture 8.9. *Let $V \subset \mathbb{P}^n$ be an irreducible projective variety of degree $d \geq 2$ defined over $\mathbb{Q}$. Then*

$$N_V(B) \ll_V B^{\dim V}(\log B)^{c_V}$$

for some constant $c_V > 0$.

This is nowadays called the *Dimension Growth Conjecture*. There has been written many papers, especially by Browning, Heath-Brown and Salberger, to find ways to prove this conjecture. A listing of references of these papers can be found in [6]. A slightly weaker version of this conjecture is now a theorem proved by Salberger [31] and is stated as follows.

Theorem 8.10. *Let $V \subset \mathbb{P}^n$ be an irreducible projective variety of degree $d \geq 2$. Then we have*

$$N_V(B) \ll_{\epsilon,V} B^{\dim V+\epsilon},$$

for any $\epsilon > 0$.

8.4 Rational points close to a manifold

In a recent paper of Huang [18], the problem of estimating

$$N_{\mathcal{M}}(Q,\eta) := \#\{\frac{\mathbf{p}}{q} \in \mathbb{Q}^n : 1 \leq q \leq Q, \mathrm{dist}(\frac{\mathbf{p}}{q}, \mathcal{M}) \leq \eta/q\}, \tag{8.4}$$

with $\mathcal{M} \subset \mathbb{R}^n$ is a bounded submanifold of dimension m, $Q > 1$ and $\eta \in (0, 1/2)$, $\mathbf{p} \in \mathbb{Z}^n$, $q \in \mathbb{Z}$ and $\mathrm{dist}(\cdot,\cdot) = \inf_{\mathbf{y}\in\mathcal{M}} |\frac{\mathbf{p}}{q} - \mathbf{y}|$, is central. We can see this problem as estimating how many points there are close to a manifold instead of points on a manifold as is being done in the Dimension Growth Conjecture (Conjecture 8.9).

Let $\mathcal{S}$ be a compact hypersurface that is given in the Monge form

$$(\mathbf{x}, f(\mathbf{x})), \quad \mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{D}, \tag{8.5}$$

where $\mathcal{D} \subset \mathbb{R}^{n-1}$ is a connected bounded open set and f a smooth function on $\mathcal{D}$. Assume that for all $\mathbf{x} \in \mathcal{D}$, the Hessian matrix $\nabla^2 f(x) := \left\{\frac{\partial^2 f}{\partial x_i \partial x_j}\right\}$ satisfies

$$0 < C_1 \leq |\det \nabla^2 f(x)| \leq C_2 \tag{8.6}$$

for some positive numbers C_1, C_2. This implies that the gradient

$$\nabla f := \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_{n-1}}\right)$$

is a diffeomorphism (an isomorphism between smooth manifolds), for a sufficiently small neighbourhood of any $\mathbf{x} \in \mathcal{D}$. We smooth the counting function (8.4) with a weight function $\omega \in C_0^\infty(\mathbb{R}^{n-1})$ with $\text{supp}(\omega) \subset \mathcal{D}$. Let

$$N_{\mathcal{S}}^{\omega}(Q, \eta) := \sum_{\substack{\mathbf{a} \in \mathbb{Z}^{n-1} \\ q \leq Q \\ ||qf(\mathbf{a}/q)|| < \eta}} \omega\left(\frac{\mathbf{a}}{q}\right) \tag{8.7}$$

be the smoothened counting function. Then Theorem 2 in [18] states the following.

Theorem 8.11. *If $\mathcal{S}$ is a hypersurface given by* (8.5)*, that satisfies* (8.6) *and $\nabla f : \mathcal{D} \to \nabla f(\mathcal{D})$ is a diffeomorphism, then*

$$N_{\mathcal{S}}^{\omega}(Q, \eta) = \frac{2\hat{\omega}(0)}{n} \eta Q^n + O_{\mathcal{S},\omega}(E_n(Q)). \tag{8.8}$$

Here $E_3(Q) = Q^2 \exp(c\sqrt{\log Q})$ and $E_n(Q) = Q^{n-1}(\log Q)^\kappa$ for $n \geq 4$, where c and κ are some positive constants.

This theorem is strongly related to the Dimension Growth Conjecture (Conjecture 8.9). In fact, this theorem is motivated by the conjecture in a certain sense. In the definition of $N_{\mathcal{S}}^{\omega}(Q, \eta)$ in (8.7), we consider η as the distance of points close to the hypersurface $\mathcal{S}$. By taking η very small, the main term will be smaller than the error term. For η arbitrarily close to zero, we can consider the error term as main term, which is $Q^{n-1}(\log Q)^\kappa$ as defined in Theorem 8.11. In this way, we are in fact counting points on the hypersurface. This is completely in line with the Dimension Growth Conjecture as the dimension of a hypersurface in $\mathbb{P}^n$ is $n-1$. A less heuristic argument of this analogue can be found in Section 7 of Huang's article [18]. We will use Theorem 8.11 in the proof of Theorem 1.2.

8.5 Bourgain versus Huang

The problem in Bourgain's research in [5], focused on finding solutions for

$$|x_1^2 + \alpha_2 x_2^2 - \alpha_3 x_3^2 - \xi| < \delta.$$

Recall that the theorem he proved is stated in Theorem 1.1. We can reformulate this problem to finding rational points near $\mathcal{C}$, by homogenizing the equation and taking $\xi = 0$. Thus, let $\mathcal{C} \subset \mathbb{R}^2$ be the planar curve given by $1 + \alpha_2 y_2^2 - \alpha_3 y_3^2 = 0$. For planar curves, a similar result as Theorem 8.11 is proved by Huang in [17]. This is labeled Theorem 3 in his article and is stated as follows.

Theorem 8.12. *Let f be a C^2 function $f : I \to \mathbb{R}$ that satisfies* (8.6)*. Furthermore, let f has a Lipschitz continuous second derivative. Then, for any $\epsilon > 0$, $0 < \eta \leq \frac{1}{2}$ and integer $Q > 1$ we have*

$$N_{\mathcal{C}}(Q, \eta) = |I|\eta Q^2 + O\left(\eta^{\frac{1}{2}}\left(\log \eta^{-1}\right) Q^{\frac{3}{2}} + Q^{1+\epsilon}\right).$$

Here the implicit constant only depends on I, C_1, C_2, ϵ and the Lipschitz constant, and is in particular independent of f, η and Q.

Note that the main term is larger than the error when $\eta \gg Q^{-2/3}$. Let us use this theorem for finding a bound for the quadratic form case, and compare this to the bound that has been found by Bourgain. We need to check if $\mathcal{C}$ satisfies the necessary conditions for Theorem 8.12. First, we have that $\det \nabla^2 f = -4\alpha_2\alpha_3$, where $f(y_2, y_3) = 1 + \alpha_2 y_2^2 - \alpha_3 y_3^2$. Then $|\det \nabla^2 f|$ is definitely bounded by a positive constant, since $\alpha_2, \alpha_3 > 0$. Since this is a constant, it is also Lipschitz continuous. Let $I = [\frac{1}{4}, 1]$, so that we can compare it to the weight functions used by Bourgain. We have $\eta = \frac{\delta}{N}$, since $|x_1| \sim N$. Therefore we have

$$N_{\mathcal{C}}(N, \delta N^{-1}) = \frac{3}{4}\delta N + O(\delta^{\frac{1}{2}}(\log(\delta^{-1}N))N^{\frac{1}{2}} + N^{1+\epsilon}) \tag{8.9}$$

for $0 < \delta \leq \frac{1}{2}N$. Hence, for $\eta \gg Q^{-2/3}$, and for $\delta N^{1/3} \to \infty$, we find at least one solution for $|x_1^2 + \alpha_2 x_2^2 - \alpha_3 x_3^2| < \delta$. In other words, we find at least one solution when

$$\min_{\substack{x \in \mathbb{Z}^3 \\ \max_i |x_i| \sim N}} |x_1^2 + \alpha_2 x_2^2 - \alpha_3 x_3^2| \ll N^{-1/3+\epsilon} \text{ for all } \epsilon > 0.$$

Although Huang found this groundbreaking result on counting rational points on planar curves, comparing this to Theorem 1.1, the bound given by Theorem 8.12 is worse than Bourgain has found whether the Lindelöf hypothesis is being assumed or not. Let us elaborate why this is the case. The conditions of Theorem 8.12 are based on a planar curve in $\mathbb{R}^2$. This is a very general statement, while the quadratic form of Bourgain is based on the fact that it is diagonalisable, which allows us to write it in the form $Q(x) = x_1^2 + \alpha_2 x_2^2 - \alpha_3 x_3^2$, which then makes it possible to split the function into two logarithms as is being done in equation (2.2) in [5]. Here α_3 is being separated and this is a crucial step in the proof, which is only possible if we are considering diagonalisable forms. This motivates our decision to follow the strategy of Bourgain in proving our main theorem.

On the other hand, Huang's theorem gives a stronger result, in the sense that the bound holds for *all* α_2 and α_3, while Bourgain's theorem is averaged over α_3, with a small exceptional set for which the bound may not hold. Therefore, if the two bounds would be equal, the theorem of Huang is preferred.

As the structure gives us ways to compute better bounds for three variables than the theorems of Huang do, we will follow Bourgain's method. However, we still use the results of Huang in our calculations, as Theorem 8.11 gives better results for higher dimensions. When working in $n = 4$, his results are good enough and will help us to compute better results.

9 The main problem

We generalise the work of Bourgain [5] and Schindler [33] and prove similar results for the inequality

$$|G_k(\mathbf{x}) - \alpha_3 x_3^l| < \delta, \tag{9.1}$$

where we recall from (1.4) that $G_k(\mathbf{x})$ is a binary form of degree k, such that

$$G_k(\mathbf{x}) = x_1^k - \alpha_{2,1}x_1x_2^{k-1} - \ldots - \alpha_{2,k}x_2^k$$

where $\alpha_{2,i} \in \mathbb{R}$, $1 \leq i \leq k$, not all equal to zero. Furthermore, let $k, l \in \mathbb{Z}$ fixed, $\alpha_3 \in \mathbb{R}_{>0}$ and $\delta > 0$. Our goal is to find non-trivial integer solutions of (9.1), where x_1, x_2 range over a box of size N^l and x_3 over a box of size N^k. Here we differ from the proofs of Bourgain and Schindler. These different sizes of x_1, x_2 and x_3 are chosen to obtain a more well-readable bound, as the exponent of N turns out to be more simplified than taking x_1, x_2, x_3 all of size N. We let δ be sufficiently large and grow with size N. We study this problem on average over $\alpha_3 \in [\frac{1}{2}, 1]$.

Note that when

$$\alpha_{2,1} = \alpha_{2,2} = \ldots = \alpha_{2,k-1} = 0, \alpha_{2,k} \neq 0,$$

we obtain results for the ternary form

$$x_1^k - \alpha_2 x_2^k - \alpha_3 x_3^l.$$

This is the ternary form as considered in Theorem 1.2. Furthermore, for k even,

$$\alpha_{2,1}, \ldots, \alpha_{2,k/2-1}, \alpha_{2,k/2+1} \ldots \alpha_{2,k-1} = 0, \;\; \alpha_{2,k/2}, \alpha_{2,k} \neq 0,$$

we obtain results for

$$x_1^k - \alpha_{2,k/2}x_1^{k/2}x_2^{k/2} - \alpha_{2,k}x_2^k,$$

which is the ternary form considered in Theorem 1.3.

The steps in the proof of Theorem 1.2 and Theorem 1.3 are as follows. We start with proving results for the general inequality (9.1). We find a lower bound for a certain counting function that will lead to the smallest solution for (9.1). We split this counting function in two parts: the main part and the error term. By showing the contribution of the error term is small, the lower bound of the main part will be our solution. We prove results for three examples of the general polynomial $G_k(\mathbf{x})$.

Let $0 < a_i < b_i$ be real parameters for $1 \leq i \leq 3$. We introduce three smooth bump functions, i.e., smooth functions $f : \mathbb{R}^n \to \mathbb{R}$ that are compactly supported. Let $0 \leq \omega_i \leq 1$, $1 \leq i \leq 3$, be smooth bump functions satisfying $\omega_i = 1$ on $[\frac{1}{2}a_i, \frac{3}{4}b_i]$ and $\text{supp}(\omega_i) \subset [\frac{1}{4}a_i, b_i]$.

Furthermore, let

$$\begin{aligned}
\left(\frac{1}{4}a_1\right)^k - \sum_{i=1}^{k-1} \alpha_{2,i} b_1^i b_2^{k-i} - \alpha_{2,k} b_2^k > 0 \\
\frac{1}{2}\left(\frac{1}{4}a_3\right)^l > \left(\frac{1}{4}a_1\right)^k - \frac{1}{4^k}\sum_{i=1}^{k-1} \alpha_{2,i} a_1^i a_2^{k-i} - \alpha_{2,k}\left(\frac{1}{4}a_2\right)^k \\
b_1^k - \sum_{i=1}^{k-1} \alpha_{2,i} b_1^i b_2^{k-i} - \alpha_{2,k} b_2^k > b_3^l
\end{aligned} \tag{9.2}$$

Roughly speaking, the first inequality is to make sure that the logarithm of $G_k(x_1, x_2)$ is well defined. The second is to bound the error term $f_4(\alpha_3)$ of $f_2(\alpha_3)$, which will be introduced in Lemma 9.1. The last inequality is to make sure that the solution set of the considered Diophantine inequality is nonempty.

Let $0 \leq \omega_0 \leq 1$ be a smooth bump function satisfying $\omega_0 = 1$ on $[-1, 1]$, supp $\omega_0 \subset [-2, 2]$ and $\omega_0(t) = \omega_0(-t)$. Furthermore, define the weight function

$$\boldsymbol{\omega}(\mathbf{x}) := \prod_{i=1}^{3} \omega_i(x_i).$$

Let $\boldsymbol{N} := (N^l, N^l, N^k)$ such that

$$\boldsymbol{\omega}\left(\frac{\mathbf{x}}{\boldsymbol{N}}\right) = \omega_1\left(\frac{x_1}{N^l}\right)\omega_2\left(\frac{x_2}{N^l}\right)\omega_3\left(\frac{x_3}{N^k}\right).$$

With this weight function, we bound x_1, x_2, x_3 in order to control their size. Consequently, we make sure system (9.2) holds.

We seek to find a lower bound for

$$\sum_{\mathbf{x} \in \mathbb{Z}^3} \boldsymbol{\omega}\left(\frac{\mathbf{x}}{\boldsymbol{N}}\right) \mathbf{1}_{[|G_k(\mathbf{x}) - \alpha_3 x_3^l| < \delta]}. \tag{9.3}$$

Let

$$|\log(G_k(\mathbf{x})) - \log(\alpha_3 x_3^l)| < \frac{\delta}{N^{kl}}$$

for $\alpha_3 \in [\frac{1}{2}, 1]$. Then if $\frac{x_i}{N_i}$ lies in the support of ω_i, and when the system (9.2) holds, we have

$$\left| e^{\log(G_k(\mathbf{x}))} - e^{\log(\alpha_3 x_3^l)} \right| \ll e^{\log(\alpha_3 x_3^l)} \frac{\delta}{N^{kl}} = \alpha_3 x_3^l \frac{\delta}{N^{kl}}.$$

Since $\alpha_3 x_3^l$ is bounded by $\alpha_3 N^{kl} \leq N^{kl}$, we see that

$$|G_k(\mathbf{x}) - \alpha_3 x_3^l| \ll_{c_1} \delta,$$

where the constant c_1 depends on α_2 (since α_2 is fixed and α_3 averaged over $[\frac{1}{2}, 1]$) and the support of $\boldsymbol{\omega}$. Therefore, instead of finding a lower bound for (9.3), we can find a lower bound for

$$f_1(\alpha_3) = \sum_{\mathbf{x} \in \mathbb{Z}^3} \boldsymbol{\omega}\left(\frac{\mathbf{x}}{N}\right) \mathbf{1}_{[|\log(G_k(\mathbf{x})) - \log(\alpha_3 x_3^l)| < \frac{\delta}{N^{kl}}]} \tag{9.4}$$

instead. Similar to the proof of Section 2 in [33], we define $T = \frac{2N^{kl}}{\delta}$. Furthermore, we define the two following functions:

$$F_1(t) := \sum_{x_1, x_2 \in \mathbb{Z}} \omega_1 \left(\frac{x_1}{N^l} \right) \omega_2 \left(\frac{x_2}{N^l} \right) e^{it \log(G_k(\mathbf{x}))} \tag{9.5}$$

$$F_2(t) := \sum_{x_3 \in \mathbb{Z}} \omega_3 \left(\frac{x_3}{N^k} \right) e^{it \log(x_3)}. \tag{9.6}$$

Furthermore, let $\widehat{\omega_0}$ be the Fourier transform of ω_0, i.e.,

$$\widehat{\omega_0}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \omega_0(t) e^{-itx} dt.$$

We obtain the following lower bound for $f_1(\alpha_3)$.

Lemma 9.1. *Let $f_1(\alpha_3)$ be as in* (9.4) *and define*

$$f_2(\alpha_3) := \frac{1}{T} \int_{\mathbb{R}} \widehat{\omega_0} \left(\frac{t}{T} \right) F_1(t) \overline{F_2(lt)} e^{-it \log(\alpha_3)} dt. \tag{9.7}$$

Then $f_1(\alpha_3) \geq f_2(\alpha_3)$.

Proof. As ω_1 is a real-valued function,

$$\begin{aligned} f_2(\alpha_3) &= \frac{1}{T} \int_{\mathbb{R}} \widehat{\omega_0} \left(\frac{t}{T} \right) F_1(t) \overline{F_2(lt)} e^{-it \log(\alpha_3)} dt \\ &= \frac{1}{T} \int_{\mathbb{R}} \widehat{\omega_0} \left(\frac{t}{T} \right) \sum_{\mathbf{x} \in \mathbb{Z}^3} \boldsymbol{\omega} \left(\frac{\mathbf{x}}{\boldsymbol{N}} \right) e^{it \log(G_k(\mathbf{x}))} \overline{e^{lit \log(x_3)}} e^{-it \log(\alpha_3)} dt \\ &= \frac{1}{T} \sum_{\mathbf{x} \in \mathbb{Z}^3} \boldsymbol{\omega} \left(\frac{\mathbf{x}}{\boldsymbol{N}} \right) \int_{\mathbb{R}} \widehat{\omega_0} \left(\frac{t}{T} \right) e^{it(\log(G_k(\mathbf{x})) - \log(\alpha_3 x_3^l))} dt. \end{aligned}$$

Write $t = Tt'$. We obtain

$$f_2(\alpha_3) = \frac{1}{T} \sum_{\mathbf{x} \in \mathbb{Z}^3} \boldsymbol{\omega} \left(\frac{\mathbf{x}}{\boldsymbol{N}} \right) \int_{\mathbb{R}} T \widehat{\omega_0}(t') e^{iTt'(\log(G_k(\mathbf{x})) - \log(\alpha_3 x_3^l))} dt'.$$

Using the inverse Fourier transform (see Section 3.1), we find

$$f_2(\alpha_3) = \sum_{\mathbf{x} \in \mathbb{Z}^3} \boldsymbol{\omega} \left(\frac{\mathbf{x}}{\boldsymbol{N}} \right) \omega_0(T(\log(G_k(\mathbf{x})) - \log(\alpha_3 x_3^l)).$$

As $\omega_0 = 1$ on the interval $[-1, 1]$, equivalently this bump function is equal to one if

$$|T(\log(G_k(\mathbf{x})) - \log(\alpha_3 x_3^l)| \leq 1,$$

which means

$$|\log(G_k(\mathbf{x})) - \log(\alpha_3 x_3^l)| \leq \frac{1}{T} = \frac{1}{2} \frac{\delta}{N^{kl}}.$$

Therefore, $f_2(\alpha_3) \leq f_1(\alpha_3)$. $\square$

Next, we split $\widehat{\omega_0}(\frac{t}{T})$ as $\widehat{\omega_0}(\frac{t}{\sqrt{N}}) + (\widehat{\omega_0}(\frac{t}{T}) - \widehat{\omega_0}(\frac{t}{\sqrt{N}}))$. We call the first contribution $f_3(\alpha_3)$ and the second $f_4(\alpha_3)$. Then

$$f_3(\alpha_3) = \frac{1}{T}\sum_{\mathbf{x}\in\mathbb{Z}^3} \boldsymbol{\omega}\left(\frac{\mathbf{x}}{\boldsymbol{N}}\right)\int_{\mathbb{R}} \widehat{\omega_0}\left(\frac{t}{\sqrt{N}}\right) e^{it(\log(G_k(\mathbf{x}))-\log(\alpha_3 x_3^l))}dt$$

and with the change of variables $t = \sqrt{N}t'$, this is equal to

$$\frac{\sqrt{N}}{T}\sum_{\mathbf{x}\in\mathbb{Z}^3} \boldsymbol{\omega}\left(\frac{\mathbf{x}}{\boldsymbol{N}}\right)\omega_0(\sqrt{N}(\log(G_k(\mathbf{x})) - \log(\alpha_3 x_3^l))).$$

Similar to the calculations in the proof of Lemma 9.1, we obtain

$$f_3(\alpha_3) \geq \frac{\sqrt{N}}{T}\sum_{\mathbf{x}\in\mathbb{Z}^3} \boldsymbol{\omega}\left(\frac{\mathbf{x}}{\boldsymbol{N}}\right)\mathbf{1}_{[|\log(G_k(\mathbf{x}))-\log(\alpha_3 x_3^l)|<N^{-1/2}]}. \tag{9.8}$$

Let us estimate this bound heuristically. First, note that (9.8) implies the lower bound

$$f_3(\alpha_3) \geq \frac{\sqrt{N}}{T}\sum_{\mathbf{x}\in\mathbb{Z}^3} \boldsymbol{\omega}\left(\frac{\mathbf{x}}{\boldsymbol{N}}\right)\mathbf{1}_{[|G_k(\mathbf{x})-\alpha_3 x_3^l|<c_2 N^{kl-1/2}]}$$

with $c_2 > 0$ small enough, depending on α_3 and the support of $\boldsymbol{\omega}$.

As $x_1, x_2 \sim N^l$, $x_3 \sim N^k$, we expect $G_k(\mathbf{x}) - \alpha_3 x_3^l \sim N^{kl}$, or equivalently we expect $G_k(\mathbf{x}) - \alpha_3 x_3^l$ to lie in the interval $[-cN^{kl}, cN^{kl}]$ for some positive constant c. We expect the number of vectors $\mathbf{x}$ that lie in the support of $\boldsymbol{\omega}$ to be of size N^{2l+k}. Therefore, the distance between two solutions of $G_k(\mathbf{x}) - \alpha_3 x_3^l$ would be of size $N^{kl}N^{-2l-k} = N^{kl-2l-k}$. A visualisation of this argument can be found in Figure 1.

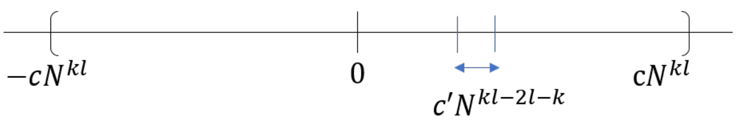

Figure 1: Visualisation of distance between two solutions of $G_k(\mathbf{x}) - \alpha_3 x_3^l$. We assume $c, c' > 0$.

The number of solutions of the counting function

$$\sum_{\mathbf{x}\in\mathbb{Z}^3} \boldsymbol{\omega}\left(\frac{\mathbf{x}}{\boldsymbol{N}}\right)\mathbf{1}_{[|G_k(\mathbf{x})-\alpha_3 x_3^l|<c_2 N^{kl-\frac{1}{2}}]}$$

therefore will be of size $N^{kl-1/2}N^{2l+k-kl} = N^{2l+k-1/2}$. Multiplying this with the factor $\sqrt{N}T^{-1}$, we expect that

$$f_3(\alpha_3) \gg N^{2l+k}T^{-1} \gg \delta N^{2l+k-kl}.$$

We can prove this lower bound using the following general lemma.

Lemma 9.2. *The number of k-th powers in the interval $(a, b] = (a, a + h]$, assuming $a > h$, is $\frac{h}{k}a^{-\frac{k-1}{k}} + O(h^2 + 1)$. In other words,*

$$\#\{m^k \in (a, a + h]\} = \frac{h}{k}a^{-\frac{k-1}{k}} + O(h^2 + 1).$$

Moreover, the minimum value of h such that $(a, a + h]$ has at least one k-th power is equal to $ka^{\frac{k-1}{k}}$.

Proof. We have

$$\begin{aligned}
\#\{m^k \in (a, b]\} &= \#\{m^k \leq b\} - \#\{m^k \leq a\} \\
&= \#\{m \leq b^{1/k}\} - \#\{m \leq a^{1/k}\} \\
&= b^{1/k} - a^{1/k} + O(1) \\
&= (a + h)^{1/k} - a^{1/k} + O(1) \\
&= a^{1/k} + \frac{h}{k}a^{-\frac{k-1}{k}} + O(h^2) - a^{1/k} + O(1) \\
&= \frac{h}{k}a^{-\frac{k-1}{k}} + O(h^2 + 1).
\end{aligned}$$

The second statement follows directly from the first. □

Fix x_1 and x_2 in the support of $\omega_1(x_1/N^l)$ and $\omega_2(x_2/N^l)$. Let $\alpha_3 \in [\frac{1}{2}, 1]$ again. In order to prove our prediction of the lower bound for $f_3(\alpha_3)$, we count the number of solutions of

$$|G_k(\mathbf{x}) - \alpha_3 x_3^l| < c_2 N^{kl-1/2},$$

which is the equivalent to counting the number of l-th powers that occur in

$$[G_k(\mathbf{x}) - c_2 N^{kl-1/2}, G_k(\mathbf{x}) + c_2 N^{kl-1/2}].$$

Assume x_3 to lie in the support of $\omega_3(x_3/N^k)$. This means x_3^l is of size N^{kl}. However, this does not imply that x_3^l has to be a kl-th power itself; it only has to be an l-th power. Using Lemma 9.2, the number of l-th powers is found to be $\gg N^{kl-1/2} \cdot N^{kl\cdot(1-l)/l} = N^{k-1/2}$. Hence there are at least $N^{k-1/2}$ choices for x_3 that satisfy the inequality. We have

$$f_3(\alpha_3) \gg \frac{\sqrt{N}}{T} N^{2l+k-1/2} \gg \delta N^{2l+k-kl}.$$

Our prediction can now be stated as lemma.

Lemma 9.3. *For every $\alpha_3 \in [\frac{1}{2}, 1]$ we have $f_3(\alpha_3) \gg \delta N^{2l+k-kl}$, where the implied constant only depends on α_2, the support of $\boldsymbol{\omega}$ and the system* (9.2).

When considering $f_3(\alpha_3)$ as the main term of the counting function $f_2(\alpha_3)$, we see, that $|G_k(\mathbf{x}) - \alpha_3 x_3^l| < \delta$ has at least one solution, if $\delta \gg N^{kl-2l-k}$.

We can see (9.8) as a counting function for the Diophantine inequality

$$\frac{\alpha_3 x_3^l}{G_k(\mathbf{x})} = 1 + O(N^{-1/2}),$$

where $x_1 \sim N^l$, $x_2 \sim N^l$ and $x_3 \sim N^k$. Note

$$|\log(G_k(\mathbf{x})) - \log(\alpha_3 x_3^l)| < N^{-1/2}$$

implies that

$$\left|\frac{\alpha_3 x_3^l}{G_k(\mathbf{x})}\right| < e^{N^{-1/2}} = 1 + N^{-1/2} + \dots$$

by writing out the Taylor expansion. We obtain

$$\begin{aligned}\alpha_3 x_3^l &= (G_k(\mathbf{x})) + O((G_k(\mathbf{x}))N^{-1/2}) \\ &= (G_k(\mathbf{x})) + O(N^{kl} \cdot N^{-1/2})\end{aligned}$$

and conclude

$$G_k(\mathbf{x}) - \alpha_3 x_3^l = O(N^{kl-1/2}).$$

The other term in the counting function, $f_4(\alpha_3)$, can be seen as an error term. We will estimate this term as a function of α_3 by

$$\textsc{Meas}_1 := \text{meas}\{\alpha_3 \in [\frac{1}{2}, 1] : |f_4(\alpha_3)| \geq (1/2)c_3\delta N^{2l+k-kl}\}. \tag{9.9}$$

Here c_3 only depends on α_2 and the support of $\boldsymbol{\omega}$. The factor $\frac{1}{2}$ is to make sure the error term will not outgrow $f_3(\alpha_3)$. Recall Chebyshev's inequality in Theorem 6.7. Take $X = [\frac{1}{2}, 1]$, $f(x) = f_3(x)$, $g(t) = t^2$, and $t = \frac{1}{2}c_3\delta N^{2l+k-kl}$. We obtain

$$\textsc{Meas}_1 \leq \frac{4N^{2kl-4l-2k}}{c_3^2\delta^2} \int_{\frac{1}{2}}^{1} |f_4(\alpha_3)|^2 d\alpha_3.$$

Remark 9.4. From here on, we assume $\delta \leq N^{kl-1/2}$. If not, then $\delta > N^{kl-1/2}$ implies $T \ll N^{kl}/N^{kl-1/2}$ which means $|G_k(\mathbf{x}) - \alpha_3 x_3^l| < 1/T$ always has a solution.

We prove three lemmas regarding the smooth bump function ω_0.

Lemma 9.5. *We have*

$$\left|\widehat{\omega_0}\left(\frac{t}{T}\right) - \widehat{\omega_0}\left(\frac{t}{\sqrt{N}}\right)\right| \leq c_4$$

for some constant $c_4 > 0$.

Proof. We note two things about ω_0. First of all, we have $|\omega_0(t)| \leq 1$ for $t \in \mathbb{R}$. Second, the support of ω_0 lies in $[-2, 2]$. Hence

$$|\widehat{\omega_0}(t)| = \int_{\mathbb{R}} \omega_0(x)e^{-itx}dx = \int_{-2}^{2} \omega_0(x)e^{-itx}dx$$

and so

$$|\widehat{\omega_0}(t)| \leq \int_{-2}^{2} |e^{-itx}|dx = c$$

for some positive constant c as the integrand and interval are bounded. Hence the difference of the two is also bounded by a constant. $\square$

Lemma 9.6. *Let t be small, say $|t| < N^{1/10}$. Then*

$$\left|\widehat{\omega_0}\left(\frac{t}{T}\right) - \widehat{\omega_0}\left(\frac{t}{\sqrt{N}}\right)\right| \ll \frac{t^2}{N}.$$

Proof. The Maclaurin expansion of $\widehat{\omega_0}\left(\frac{t}{T}\right)$ is

$$\widehat{\omega_0}\left(\frac{t}{T}\right) = \widehat{\omega_0}(0) + \widehat{\omega_0}'(0)\frac{t}{T} + \frac{1}{2}\widehat{\omega_0}''(0)\left(\frac{t}{T}\right)^2 + \dots$$

Note that the first derivative of ω_0 is equal to zero for $t = 0$, as $\omega_0 = 1$ at $[-1, 1]$. As $\widehat{\omega}_0$ is concave up, the second derivative is positive. We obtain

$$\begin{aligned}\left|\widehat{\omega_0}\left(\frac{t}{T}\right) - \widehat{\omega_0}\left(\frac{t}{\sqrt{N}}\right)\right| &\leq \left|\widehat{\omega_0}\left(\frac{t}{T}\right) - 1\right| - \left|\widehat{\omega_0}\left(\frac{t}{\sqrt{N}}\right) - 1\right| \\ &\ll \left(\frac{t}{T}\right)^2 - \left(\frac{t}{\sqrt{N}}\right)^2 \ll \left(\frac{t}{T}\right)^2\end{aligned}$$

The last inequality follows as t is small and $\delta \leq N^{kl-1/2}$. □

Lemma 9.7. *For t large, say $|t| > N^{1/10}$, we have*

$$\left|\widehat{\omega_0}\left(\frac{t}{T}\right) - \widehat{\omega_0}\left(\frac{t}{\sqrt{N}}\right)\right| \ll \left(\frac{T}{|t|}\right)^{10}.$$

Here the power 10 is an arbitrary value, i.e., one could take any value $N \geq 0$.

This is a result of the following theorem on oscillatory integrals [37].

Theorem 9.8. *Let*

$$I(\lambda) = \int_a^b e^{i\lambda\phi(x)}\psi(x)dx,$$

where ϕ is a real valued smooth function, ψ a complex-valued and smooth function and with compact support in (a, b). Let $\phi'(x) \neq 0$ for all $x \in [a, b]$. Then

$$I(\lambda) = O(\lambda^{-N}) \text{ as } \lambda \to \infty$$

for all $N \geq 0$.

Proof. This proof uses an integration by part argument. Note $(e^{i\lambda\phi(x)})' = (i\lambda\phi'(x))e^{i\lambda\phi(x)}$. We define the differential operator D on a function f as

$$Df(x) = (i\phi'(x))^{-1}\frac{df}{dx}.$$

Then $\lambda^{-1}D(e^{i\lambda\phi(x)}) = e^{i\lambda\phi(x)}$. Since $\phi'(x) \neq 0$ by assumption, this differential operator D is well-defined. Since the differential operator is antisymmetric, its transpose (denoted by ${}^tD$) is given by

$${}^tDf(x) = \frac{-d}{dx}\left(\frac{f}{i\phi'(x)}\right)$$

Integration by parts leads to

$$\begin{aligned}\int_a^b e^{i\lambda\phi(x)}\psi(x)dx &= \int_a^b D(e^{i\lambda\phi(x)})\psi(x)dx\\ &= \frac{e^{i\lambda\phi(x)}\psi(x)}{i\lambda\phi'(x)}\Big|_a^b + \int_a^b \lambda^{-1}e^{i\lambda\phi(x)}({}^tD)(\psi(x))dx.\end{aligned}$$

By the support of $\psi(x)$, the first term on the right hand side is equal to zero. Repeating this argument N times, we find

$$I(\lambda) = \lambda^{-N}\int_a^b \lambda^{-1}e^{i\lambda\phi(x)}({}^tD)^N(\psi(x))dx.$$

Since the absolute value of the resulting integral is finite, we find

$$|I(\lambda)| \ll |\lambda|^{-N}.$$

□

We return to the proof of Lemma 9.7.

Proof of Lemma 9.7. Using Theorem 9.8 with $[a,b] = [-2,2]$, $\psi(x) = \omega_0(x)$ and $\phi(x) = -x$, we see

$$\widehat{\omega_0}(t) = \int_{-2}^2 \omega_0(x)e^{-itx}dx$$

is an oscillatory integral. Also $\phi'(x) = -1$. Therefore, $|\widehat{\omega_0}(t)| \ll |t|^{-N}$, and since t is large,

$$\left|\widehat{\omega_0}\left(\frac{t}{T}\right) - \widehat{\omega_0}\left(\frac{t}{\sqrt{N}}\right)\right| \le \left|\widehat{\omega_0}\left(\frac{t}{T}\right)\right| \ll \left(\frac{T}{|t|}\right)^{10}.$$

□

We summarise these three lemmas by

$$\left|\widehat{\omega_0}\left(\frac{t}{T}\right) - \widehat{\omega_0}\left(\frac{t}{\sqrt{N}}\right)\right| \le c_5 \min\left(1, \frac{t^2}{N}, \left(\frac{T}{|t|}\right)^{10}\right). \tag{9.10}$$

Here c_5 is a positive constant depending on N and the support of ω_0. Let

$$I_1 := \int_{1/2}^1 |f_4(\alpha_3)|^2 d\alpha_3.$$

We find

$$\begin{aligned}|f_4(\alpha_3)|^2 &= \left|\frac{1}{T}\int_{\mathbb{R}}\left(\widehat{\omega_0}\left(\frac{t}{T}\right) - \widehat{\omega_0}\left(\frac{t}{\sqrt{N}}\right)\right)F_1(t)\overline{F_2(lt)}e^{-it\log(\alpha_3)}dt\right|^2\\ &\le \frac{1}{T^2}\int_R \left|\widehat{\omega_0}\left(\frac{t}{T}\right) - \widehat{\omega_0}\left(\frac{t}{\sqrt{N}}\right)\right|^2 |F_1(t)|^2|F_2(lt)|^2|e^{-it\log(\alpha_3)}|^2dt\\ &\le \frac{c_5^2}{T^2}\int_{\mathbb{R}}\min\left(1, \frac{t^4}{N^2}, \left(\frac{T}{|t|}\right)^{20}\right)|F_1(t)|^2|F_2(lt)|^2|e^{-it\log(\alpha_3)}|^2dt.\end{aligned}$$

With change of variables $\beta = \log(\alpha_3)$,

$$I_1 = \int_{\frac{1}{2}}^{1} |f_4(\alpha_3)|^2 d\alpha_3 = \int_0^{|\log(\frac{1}{2})|} |f_4(e^\beta)|^2 d\beta.$$

Let

$$H(t) := \frac{c_5}{T} \min\left(1, \frac{t^2}{N}, \left(\frac{T}{|t|}\right)^{10}\right) F_1(t)\overline{F_2(lt)}.$$

Recall Parseval's identity as stated in Theorem 3.6. Using this, we obtain

$$I_1 \leq \int_{\mathbb{R}} \int_{\mathbb{R}} |H(t)e^{-it\beta} d\beta|^2 dt = \int_{\mathbb{R}} |\widehat{H}(t)|^2 dt = \int_{\mathbb{R}} |H(t)|^2 dt.$$

Hence

$$I_1 \leq \frac{c_5^2}{T^2} \int_{\mathbb{R}} \min\left(1, \frac{t^4}{N^2}, \left(\frac{T}{|t|}\right)^{20}\right) |F_1(t)|^2 |F_2(lt)|^2 dt.$$

For sufficiently small t, say $|t| \leq N^{1/10}$, the minimum will be

$$\min\left(1, \frac{t^4}{N^2}, \left(\frac{T}{|t|}\right)^{20}\right) = \frac{t^4}{N^2}.$$

However, for $|t| > N^{1/10}$, this term will never be the minimum. Therefore, we split the integral and obtain

$$I_1 \leq \frac{c_5^2}{T^2}\left(\int_{|t|\leq N^{1/10}} \frac{t^4}{N^2} |F_1(t)|^2 |F_2(lt)|^2 dt + \int_{|t|>N^{1/10}} \min\left(1, \left(\frac{T}{|t|}\right)^{20}\right) |F_1(t)|^2 |F_2(lt)|^2 dt\right).$$

Note that for small t, we can bound $|F_1(t)|^2$ by N^{4l} and $|F_2(lt)|^2$ by N^{2k}. In this way we find

$$I_1 \leq c_6 \delta^2 N^{-3/2-2kl+4l+2k} + \frac{c_5^2}{T^2} \int_{|t|>N^{1/10}} \min\left(1, \left(\frac{T}{|t|}\right)^{20}\right) |F_1(t)|^2 |F_2(lt)|^2 dt.$$

Define I_2 as

$$\begin{aligned} I_2 &:= \frac{4N^{2kl-4l-2k}}{c_3^2 \delta^2} \cdot \frac{c_5^2}{T^2} \int_{|t|>N^{1/10}} \min\left(1, \left(\frac{T}{|t|}\right)^{20}\right) |F_1(t)|^2 |F_2(lt)|^2 dt \\ &= \frac{c_5^2}{c_3^2} N^{-4l-2k} \int_{|t|>N^{1/10}} \min\left(1, \left(\frac{T}{|t|}\right)^{20}\right) |F_1(t)|^2 |F_2(lt)|^2 dt. \end{aligned} \tag{9.11}$$

We find

$$\text{Meas}_1 \leq c_7 N^{-3/2} + I_2.$$

for some constant $c_7 > 0$. Since $c_7 N^{-3/2}$ is already small enough for the purposes of this estimation, we focus on estimating I_2. Define

$$I_3 := \int_{\mathbb{R}} \min\left(1, \left(\frac{T}{|t|}\right)^{10}\right) |F_1(t)|^2 dt.$$

Then I_2 is bounded by

$$I_2 \ll N^{-4l-2k} \max_{|t|>N^{1/10}} \left(\min\left(1, \left(\frac{T}{|t|}\right)\right) |F_2(t)| \right)^2 I_3. \tag{9.12}$$

Furthermore, let

$$I(y) := \int_{\mathbb{R}} \min\left(1, \left(\frac{T}{|t|}\right)^{10}\right) e^{ity} dt.$$

This can be bounded by

$$|I(y)| \le \int_{\mathbb{R}} \min\left(1, \left(\frac{T}{|t|}\right)^{10}\right) dt \le 2T + \int_{|t|>T} \left(\frac{T}{|t|}\right)^{10} dt \ll T.$$

Furthermore, for $y \neq 0$, Theorem 9.8 implies

$$I(y) = \int_{|t|\le T} e^{ity} + \int_{|t|>T} \left(\frac{T}{|t|}\right)^{10} e^{ity} dt \ll \frac{1}{|y|}.$$

We conclude

$$|I(y)| \ll \min\left(\frac{1}{|y|}, T\right). \tag{9.13}$$

We continue on bounding I_3. Let $\mathbf{y} = (y_1, \dots, y_4)$ and let

$$\tilde{\boldsymbol{\omega}}(\mathbf{y}) := \omega_1(y_1)\omega_2(y_2)\omega_1(y_3)\omega_2(y_4).$$

Then

$$\begin{aligned} I_3 &= \int_{\mathbb{R}} \min\left(1, \left(\frac{T}{|t|}\right)^{10}\right) \left| \sum_{x_1, x_2 \in \mathbb{Z}} \omega_1\left(\frac{x_1}{N^l}\right) \omega_2\left(\frac{x_2}{N^l}\right) e^{it\log(G_k(\mathbf{x}))} \right|^2 dt \\ &= \int_{\mathbb{R}} \min\left(1, \left(\frac{T}{|t|}\right)^{10}\right) \sum_{\mathbf{y}\in\mathbb{Z}^4} \tilde{\boldsymbol{\omega}}\left(\frac{\mathbf{y}}{N^l}\right) e^{it(\log(G_k(y_1,y_2)) - \log(G_k(y_3,y_4)))} dt \\ &\ll \sum_{\mathbf{y}\in\mathbb{Z}^4} \tilde{\boldsymbol{\omega}}\left(\frac{\mathbf{y}}{N^l}\right) \min\left(\frac{1}{|\log(G_k(y_1,y_2)) - \log(G_k(y_3,y_4))|}, T\right). \end{aligned}$$

Here (9.13) is used in the last inequality. By the system (9.2), both logarithms are real, i.e., $G_k(y_1, y_2) > 0$ and $G_k(y_3, y_4) > 0$. By $\tilde{\boldsymbol{\omega}}$, the values for y_1, y_2 are bounded above by b_1, b_2 respectively and bounded from below by $\frac{1}{4}a_1, \frac{1}{4}a_2$ respectively. Hence the smallest possible value for $G_k(y_1, y_2)$ is at least

$$\left(\frac{1}{4}a_1\right)^k - \sum_{i=1}^{k-1} \alpha_{2,i} b_1^i b_2^{k-1} - \alpha_{2,k} b_2^k,$$

which is positive by the first inequality in (9.2). The same holds for $G_k(y_3, y_4)$.

Furthermore, regarding the support of $\tilde{\boldsymbol{\omega}}$, each y_i, $i = 1, \dots, 4$, is bounded by a constant times N^l. Hence

$$|\log(G_k(y_1, y_2)) - \log(G_k(y_3, y_4))| \ll \log N.$$

We define

$$I_4 := U \sum_{\mathbf{y}\in\mathbb{Z}^4} \tilde{\boldsymbol{\omega}}\left(\frac{\mathbf{y}}{N^l}\right) \mathbf{1}_{[|\log(G_k(y_1,y_2))-\log(G_k(y_3,y_4))|<\frac{1}{U}]} \tag{9.14}$$

for $\frac{1}{\log N} \ll U \leq T$. Define

$$\Delta := \log(G_k(y_1,y_2)) - \log(G_k(y_3,y_4)).$$

We split I_3 in two parts; $\Delta < \frac{1}{T}$ and $\Delta \geq \frac{1}{T}$. For the latter part, we use dyadic decomposition in order to find an expression for it. Let $k \geq 0$. We find

$$\begin{aligned} I_3 &\ll \sum_{\mathbf{y}\in\mathbb{Z}^4} \tilde{\boldsymbol{\omega}}\left(\frac{\mathbf{y}}{N^l}\right) \min\left(\frac{1}{|\Delta|}, T\right) \\ &= T\sum_{\mathbf{y}\in\mathbb{Z}^4} \tilde{\boldsymbol{\omega}}\left(\frac{\mathbf{y}}{N^l}\right) \mathbf{1}_{|\Delta|<\frac{1}{T}} + \sum_{U=2^{-k}T} U \sum_{\substack{\mathbf{y}\in\mathbb{Z}^4 \\ \frac{1}{2U}<\Delta<\frac{1}{U}}} \tilde{\boldsymbol{\omega}}\left(\frac{\mathbf{y}}{N^l}\right) \mathbf{1}_{|\Delta|<\frac{1}{U}} \\ &= I_4(T) + \sum_{U=2^{-k}T} I_4(U). \end{aligned}$$

Note $\frac{1}{\log N} \ll 2^{-k}T \leq T$ is equivalent to $2^k < T\log N \ll N^\alpha$ for some $\alpha > 0$. Hence

$$I_3 \ll \log N \sup_{\frac{1}{\log N}\ll U\leq T} I_4(U).$$

Furthermore, I_4 can be bounded by

$$I_4(U) \leq U \sum_{\mathbf{y}\in\mathbb{Z}^4} \tilde{\boldsymbol{\omega}}\left(\frac{\mathbf{y}}{N^l}\right) \mathbf{1}_{|G_k(y_1,y_2)-G_k(y_3,y_4)|\ll N^{kl}/U} \tag{9.15}$$

by the same reasoning as for $f_1(\alpha_3)$ in the beginning of this section.

We would like to bound the last expression. One way is to express y_1 in terms of an interval which concerns y_2, y_3, y_4. In this way, we can try to find a bound for an expression in the latter three variables, which gives us more insight in the bound.

Define

$$\mathcal{G}(\mathbf{y}) := G_k(y_1,y_2) - G_k(y_3,y_4). \tag{9.16}$$

Assume there exists a solution $\mathbf{y}^0 = (y_1^0, y_2^0, y_3^0, y_4^0)$ for

$$|\mathcal{G}(\mathbf{y})| \ll \frac{N^{kl}}{U},$$

say

$$\mathcal{G}(\mathbf{y^0}) = \epsilon$$

for some small $\epsilon > 0$. As already mentioned in Section 7, in order to use the Implicit Function Theorem, we can write $\mathcal{G}(\mathbf{y^0}) - \epsilon = 0$ and continue with the latter equality. For simplicity, we assume $\mathcal{G}(\mathbf{y^0}) = 0$.

Since G_k is a binary form, $\mathcal{G}(\mathbf{y})$ is continuously differentiable in an open set containing $\mathbf{y}^0$. By assumption, $\det(D_4(\mathcal{G}(\mathbf{y})) \neq 0$. Hence by the Implicit Function Theorem 7.1, there exists an open set $A \subset \mathbb{R}^3$ containing (y_2^0, y_3^0, y_4^0) and an open set $B \subset \mathbb{R}$ containing y_1^0, such that for each $(y_2, y_3, y_4) \in A$ there exists a unique function $\Phi(y_2, y_3, y_4) = y_1 \in B$, for which we have

$$\mathcal{G}(\Phi(y_2, y_3, y_4), y_2, y_3, y_4) = 0.$$

This function is also differentiable.

In conclusion, by the Implicit Function Theorem, we can express y_1 as a function of y_2, y_3, y_4 if the derivative is not equal to zero. However, in order to use the method introduced by Schindler [33], we need to make sure that this function is smooth on a connected open bounded set $\mathcal{D}$, and which has nonzero Hessian for all $\mathbf{y} \in \mathcal{D}$. A natural question arises: when is the Hessian of such a function nonzero? How can we decide whether we can express y_1 as a function of y_2, y_3, y_4 with nonzero Hessian in a certain domain?

We don't have an immediate answer to these questions. Luckily, there are some examples of $\mathcal{G}(\mathbf{y})$, for which it is not unknown how to express y_1 in terms of y_2, y_3, y_4. In the next subsections, we explore three examples. With these examples, we will get more acquainted with $\mathcal{G}(\mathbf{y})$ and find explicit bounds for the expression $I_4(U)$ in (9.15).

9.1 Proof of Theorem 1.2

We start with finding an upper bound of $I_4(U)$ for the inequality

$$|x_1^k - \alpha_2 x_2^k - \alpha_3 x^l| < \delta, \tag{9.17}$$

with $\alpha_2 \in \mathbb{R}, l, k \in \mathbb{Z}, \alpha_3 \in \mathbb{R}_{>0}$ and $\delta > 0$. This inequality is equivalent to equation (9.1) with $G_k(\mathbf{x}) = x_1^k - \alpha_2 x_2^k$. We follow the work of Schindler [33], who uses recent work of Huang [18]. We rewrite inequality (9.15) in terms of the inequality (9.17) and find

$$I_4(U) \leq U \sum_{\mathbf{y} \in \mathbb{Z}^4} \tilde{\boldsymbol{\omega}} \left(\frac{\mathbf{y}}{N^l} \right) \mathbf{1}_{|y_1^k - \alpha_2 y_2^k + y_3^k - \alpha_2 y_4^k| \ll N^{kl}/U}.$$

We distinguish two different cases; small and large values of U. For small values of U, say $U \ll N^l$, we fix y_2, y_3 and y_4 in the support of ω_1. Using Lemma 9.2, the number of choices for y_1 will be $\ll \frac{N^{kl}}{U} (N^{kl})^{-(k-1)/k} = \frac{N^l}{U}$. Therefore,

$$I_4(U) \ll U N^{3l} \frac{N^l}{U} \ll N^{4l}. \tag{9.18}$$

For large values of U, so for $c_7 N^l \leq U \leq T$, where c_7 is a sufficiently large positive constant. We apply a theorem of Huang [18]. Recall the notation and theorems from Section 8. Adjusting Section 3 of Schindler's work [33] to our assumptions and findings, we find the following bound for $I_4(U)$, considering c_7 sufficiently large.

$$I_4(U) \leq U \sum_{\substack{(y_2, y_3, y_4) \in \mathbb{Z}^3 \\ \frac{1}{4} a_1 N^l \leq y_3 \leq b_1 N^l \\ ||(y_3^k + \alpha_2 (y_2^k - y_4^k))^{1/k}|| \ll N^l/U}} \omega_2 \left(\frac{y_2}{N^l} \right) \omega_2 \left(\frac{y_4}{N^l} \right). \tag{9.19}$$

Then, by defining the function

$$f := (1 + \alpha_2 z_2^k - \alpha_2 z_4^k)^{1/k}$$

for $z_2, z_4 \in \mathcal{D}$ where $\mathcal{D} := [(1/4)a_2 b_1^{-1}, 4b_2 a_1^{-1}]^2$, Schindler showed that the Hessian of f is strictly negative on $\mathcal{D}$. She defined a new smooth weight function $0 \leq \omega_4 \leq 1$ such that $\mathrm{supp}(\omega_4) \subset [(1/4)a_2 b_1^{-1}, 4b_2 a_1^{-1}]$. We adopt this definition and find

$$I_4(U) \leq U \sum_{\substack{(y_2, y_4) \in \mathbb{Z}^2 \\ 1 \leq y_3 \leq b_1 N^l \\ ||y_3 f(y_2/y_3, y_4/y_3)|| \ll \frac{N^l}{U}}} \omega_4\left(\frac{y_2}{y_3}\right) \omega_4\left(\frac{y_4}{y_3}\right).$$

We can use Theorem 8.11 now, with $\eta \ll N^l/U$, so that $\eta < 1/2$ when c_7 is sufficiently large, and take $Q = b_1 N^l$. Deduce that

$$I_4(U) \ll U(\eta N^{3l} + N^{2l+\epsilon}). \tag{9.20}$$

As

$$I_3 \ll \log N \sup_{\frac{1}{\log N} \ll U \leq T} I_4(U),$$

we obtain a bound for I_3 which is given in the following proposition.

Proposition 9.9. *Assume that* (9.2) *holds. Then we have*

$$I_3 \ll_\epsilon N^{4l+\epsilon} + \frac{N^{kl+2l+\epsilon}}{\delta}$$

for any $\epsilon > 0$.

We return to the bound on I_2 in (9.12). We seek to obtain a bound for $F_2(t)$. Recall the definition of the Mellin transform given in Section 3. Let

$$\breve{\omega}_3(s) = \int_0^\infty \omega_3(x) x^{s-1} dx$$

be the Mellin transform of ω_3. We write

$$F_2(t) = \sum_{n \in \mathbb{Z}} \omega_3\left(\frac{n}{N^k}\right) e^{it \log n} = \sum_{n \in \mathbb{Z}} \omega_3\left(\frac{n}{N^k}\right) n^{it}.$$

Letting $a_n = n^{it}$, and

$$\xi(s) = \sum_{n \geq 1} a_n n^{-s} = \sum_{n \geq 1} n^{it-s} = \zeta(s - it)$$

where $\zeta(s)$ is the Riemann zeta function, we use Theorem 3.10 with $\phi(x) = \omega_3(x/N^k)$. We find

$$\begin{aligned} F_2(t) &= \sum_{n \geq 1} n^{it} \omega_3\left(\frac{n}{N^k}\right) \\ &= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \xi(s) \breve{\omega}_3(s) N^{ks} ds \\ &= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \zeta(s - it) \breve{\omega}_3(s) N^{ks} ds, \end{aligned}$$

provided that $\xi(s)$ converges absolutely for $\mathrm{Re}(s) = c$. We know from Section 5 that $\zeta(s)$ converges absolutely for $\mathrm{Re}(s) > 1$, so if we let $c = 2$ we obtain

$$F_2(t) = \frac{1}{2\pi i} \int_{2-i\infty}^{2+i\infty} \zeta(s - it)\breve{\omega}_3(s)N^{ks}ds. \tag{9.21}$$

Write $s = \sigma + it$. We rewrite $\breve{\omega}_3$ as

$$\begin{aligned} \breve{\omega}_3(\sigma + it) &= \int_0^\infty \omega_3(x)x^{\sigma+it-1}dx \\ &= \int_0^\infty \omega_3(x)x^{\sigma-1}e^{it\log x}dx. \end{aligned}$$

By letting $\lambda = t$, $\phi(x) = \log x$, which is real in the support of ω_3, and $\psi(x) = \omega_3(x)x^{\sigma-1}$ which has compact support in $(\frac{1}{4}a_3, b_3)$, we can apply Theorem 9.8. Hence

$$\breve{\omega}_3 = O(t^{-N})$$

for all $N \geq 0$, as $t \to \infty$. By taking N large enough, we see that on vertical lines (i.e., letting $t \to \infty$), $\breve{\omega}_3$ has rapid decay. Now we can bound the contour integral in (9.21) by iT and shift the line of integration to Re $s = \frac{1}{2}$. Inside this contour, we find one pole at $1 + it$ from ζ. A visualisation of this contour is given in Figure 2.

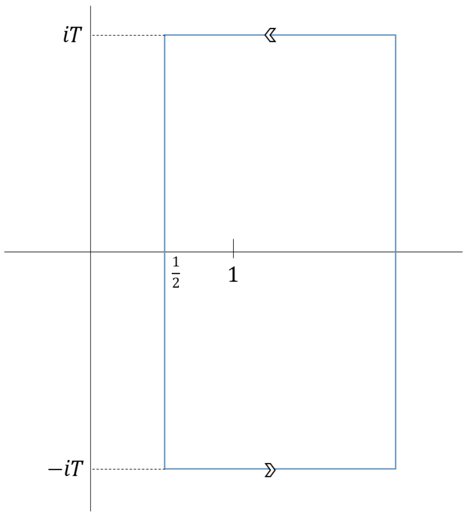

Figure 2: Contour of integration

From Section 4, we know the simple pole of ζ has residue 1. This gives a contribution

$$\breve{\omega}_3(1 + it)N^{k(1+it)}$$

to $F_2(t)$, but by the rapid decay of $\breve{\omega}_3$, this is negligible for $|t| > N^{1/10}$. Therefore, letting $s = \sigma + iy$ we can bound $F_2(t)$ by

$$F_2(t) \ll (N^k)^{1/2} \int_{-\infty}^{\infty} \frac{|\zeta(\frac{1}{2} + i(y - t))|}{1 + |y|^{10}} dy. \tag{9.22}$$

Here the denominator comes from taking $N = 10$ in applying Theorem 9.8 for $\tilde{\omega}_3$ for large values of y, and bounding by 1 for small values of y. Hence, assuming the Lindelöf hypothesis (see Section 5.2) this is bounded by $\ll (N^k)^{1/2}(1+|t|)^\epsilon$. Therefore,

$$I_2 \ll N^{-4l-2k} N^{k+\epsilon} \left(N^{4l+\epsilon} + \frac{N^{kl+2l+\epsilon}}{\delta} \right) = N^{-k+\epsilon} + \frac{N^{kl-2l-k+\epsilon}}{\delta}.$$

Returning to $\textsc{Meas}_1$, we see

$$\textsc{Meas}_1 \ll N^{-3/2} + N^{-k+\epsilon} + \frac{N^{kl-2l-k+\epsilon}}{\delta}.$$

In conclusion, we have proved the following.

Proposition 9.10. *Let $\alpha_2 > 0$ and $k, l \in \mathbb{Z}$, where $k \neq l$, be fixed. Let N be a large real parameter and assume $\delta > N^{kl-2l-k}$. Assuming the Lindelöf hypothesis for the Riemann zeta function, the inequality*

$$|x_1^k - \alpha_2 x_2^k - \alpha_3 x_3^l| < \delta$$

has a non-trivial solution where $|x_1|, |x_2| \sim N^l$, $|x_3| \sim N^k$, for all $\alpha_3 \in [1/2, 1]$, excluding an exceptional set of measure at most

$$\ll N^{-k+\epsilon} + \frac{N^{kl-2l-k+\epsilon}}{\delta}.$$

This leads to Theorem 1.2 as follows. Let $\delta > N^{kl-2l-k+\epsilon}$. Then

$$N^{-k+\epsilon} + \frac{N^{kl-2l-k+\epsilon}}{\delta} < N^{-k+\epsilon} + 1.$$

By letting $N = 2^\nu$, $\nu \in \mathbb{N}$ large enough, we obtain

$$\sum_{N=2^\nu} N^{-k+\epsilon} + \frac{N^{kl-2l-k+\epsilon}}{\delta} < \sum_{N=2^\nu} N^{-k+\epsilon} + 1 < \infty$$

as $k \geq 3$ and $2^{\nu\epsilon}$ converges. We are now in the position to apply the Borel-Cantelli Lemma 6.9 for a discrete countable collection $\{A_\nu\}_{\nu \in \mathbb{N}}$, where

$$A_\nu = (2^\nu)^{-k+\epsilon} + (2^\nu)^{kl-2l-k+\epsilon}\delta^{-1}.$$

Then for almost all $\alpha_3 \in [\frac{1}{2}, 1]$ the inequality

$$|x_1^k - \alpha_2 x_2^k - \alpha_3 x_3^l| < (2^\nu)^{kl-2l-k+\epsilon}$$

holds. In order to be able to apply the lemma for all N, we let N range from $2^\nu < N \leq 2^{\nu+1}$. If the inequality holds for 2^ν, it certainly holds for $2^\nu < N \leq 2^{\nu+1}$ as $kl - 2l - k > 0$. Therefore, we conclude that for almost all $\alpha_3 \in [\frac{1}{2}, 1]$ and $\delta = N^{kl-2l-k+\epsilon}$ we have

$$|x_1^k - \alpha_2 x_2^k - \alpha_3 x_3^l| < N^{kl-2l-k+\epsilon}.$$

This is exactly part *(i)* in Theorem 1.2.

9.1.1 Unconditional case

As the Lindelöf Hypothesis is not proved yet, some things about the unconditional case need to be mentioned. In this case, we have $|\zeta(\frac{1}{2}+it)| \ll (1+|t|)^{\frac{1}{6}}$, see Section 5. We obtain the bound $F_2(t) \ll N^{\frac{1}{2}k}|t|^{1/6}$. We find

$$\begin{aligned} I_2 &\ll N^{-4l-2k}N^k\left(\frac{N^{kl}}{\delta}\right)^{\frac{1}{3}}\left(N^{4l+\epsilon}+\frac{N^{kl+2l+\epsilon}}{\delta}\right) \\ &\ll N^{\frac{1}{3}kl-k+\epsilon}\delta^{-\frac{1}{3}}+N^{\frac{4}{3}kl-2l-k+\epsilon}\delta^{-\frac{4}{3}}. \end{aligned}$$

Hence

$$\text{MEAS}_1 \ll N^{-\frac{3}{2}}+N^{\frac{1}{3}kl-k+\epsilon}\delta^{-\frac{1}{3}}+N^{\frac{4}{3}kl-2l-k+\epsilon}\delta^{-\frac{4}{3}}.$$

This leads to the following proposition.

Proposition 9.11. *Let $\alpha_2 > 0$ and $k,l \in \mathbb{Z}$, where $k \neq l$, be fixed. Let N be a large real parameter and assume $N^{kl-2l-k} < \delta \leq N^{kl-1/2}$. Unconditionally, the inequality*

$$|x_1^k - \alpha_2 x_2^k - \alpha_3 x_3^l| < \delta$$

has a non-trivial solution where $|x_1|,|x_2| \sim N^l$, $|x_3| \sim N^k$, for all $\alpha_3 \in [1/2,1]$, excluding an exceptional set of measure at most

$$\ll N^{\frac{1}{3}kl-k+\epsilon}\delta^{-\frac{1}{3}}+N^{\frac{4}{3}kl-2l-k+\epsilon}\delta^{-\frac{4}{3}}.$$

Using the Borel-Cantelli Lemma 6.9 and dyadic composition would lead to a weaker bound than the one in Theorem 1.2. For a stronger bound, we use the ideas of Bourgain in [5], Section 3 and Schindler in [33], Section 4. As we assumed the size of x_1, x_2 to be different than in [33], we rephrase Lemma 4.2 and 4.3 of the Note in order to use it for our purposes. Then Lemma 4.2 becomes

Lemma 9.12. *Let $F_1(t)$ be defined as in* (9.5) *and let $T > N^{2l}$. Assume that equation* (9.2) *holds. Then*

$$\text{MEAS}\{|t| \leq T : |F_1(t)| > \lambda\} \ll TN^{2l+\epsilon}\lambda^{-2}.$$

Going through the proof of Lemma 4.2 in [33] should convince the reader that this lemma holds in our case as well. The second lemma, wich is Lemma 4.3 in [33], is rephrased as follows.

Lemma 9.13. *Assume that $|t| > N^{2l}$ and assume the system* (9.2) *holds. Then*

$$|F_1(t)| \ll N^l|t|^{1/3+\epsilon}.$$

The proof uses ideas of Titchmarsh [42] on bounding partial sums of the Epstein zeta function. This strategy uses Weyl differencing (see Section 6) and second derivative tests. As we defined our weight functions and the system of inequalities (9.2) differently from [33], we rewrite the sum in the proof of Lemma 4.3 as

$$S(X_1,X_2) := \sum_{\substack{(\frac{1}{4}a_1)^k N^l \leq x_1 < X_1 \\ (\frac{1}{4}a_2)^k N^l \leq x_2 < X_2}} e^{it\log(x_1^k-\alpha_2 x_2^k)} \ll N^l|t|^{\frac{1}{3}+\epsilon}.$$

By the first equation in the system (9.2), we have $x_1^k - \alpha_2 x_2^k \sim N^{kl}$. By the same arguments as in the proof of Lemma 4.3 of [33], we find $|t| < N^{3l}$ and $(X_1 - (\frac{1}{4}a_1)^k N^l)(X_2 - (\frac{1}{4}a_2)^k N^l) \gg N^l |t|^{\frac{1}{3}+\epsilon}$. Set $\rho := \lfloor N^l |t|^{-\frac{1}{3}} \rfloor$. As we are looking at the same binary form of degree k, all properties for f, g and G in the proof of the lemma will hold for our case as well. We find

$$S(X_1, X_2) \ll \frac{N^{2l}}{\rho} + \frac{N^{2l}}{\rho} \left(\sum_{\lambda=1}^{\rho-1} \frac{\rho^2 t^{1+\epsilon}}{N^l} \right)^{\frac{1}{2}} \ll \frac{N^{2l}}{\rho} + N^{l/2} |t|^{1+\epsilon} \rho^{1/2} \ll N^l |t|^{\frac{1}{3}+\epsilon}.$$

Here $\lambda = \max\{\mu, \nu\}$ and μ, ν were used as variables for the Weyl differencing. We use these two lemmas to prove the following proposition.

Proposition 9.14. *Let $\alpha_2 > 0$ and $l, k \geq 3$, where $k \neq l$, be fixed. Let N be a large real parameter and assume $N^{kl-2l-k} < \delta \leq N^{kl-1/2}$. The inequality*

$$|x_1^k - \alpha_2 x_2^k - \alpha_3 x_3^l| < \delta$$

has a non-trivial solution where $|x_1|, |x_2| \sim N^l$, $|x_3| \sim N^k$, for all $\alpha_3 \in [1/2, 1]$, excluding an exceptional set of measure at most

$$\ll N^{kl-k-\frac{4}{3}l+\epsilon}\delta^{-1} + N^{\frac{10}{9}kl-\frac{2}{3}l-2k+\epsilon}\delta^{-\frac{10}{9}} + N^{\frac{5}{6}kl-2k+\epsilon}\delta^{-\frac{5}{6}}.$$

One more lemma is needed for the proof. We state the lemma without proof.

Lemma 9.15 (Lemma 1 in [5])**.** *Consider the Dirichlet polynomial*

$$S(t) = \sum_{n \sim N} a_n n^{it} \text{ with } |a_n| \leq 1.$$

Then, for $T > N$,

$$\textsc{Meas}[|t| < T; |S(t)| > V] \ll N^\epsilon (N^2 V^{-2} + N^4 V^{-6} T).$$

Using these three lemmas and the idea of Section 4 in [5], we are in the position to prove the proposition.

Proof of Proposition 9.14. Recall the definition of I_2 in (9.11),

$$I_2 = \frac{c_4^2}{c_3^2} N^{-4l-2k} \int_{|t|>N^{1/10}} \min\left(1, \left(\frac{T}{|t|}\right)^{20}\right) |F_1(t)|^2 |F_2(lt)|^2 dt.$$

We subdivide the integral into

$$\int_{|t|>N^{1/10}} = \int_{N^{1/10} \leq |t| \leq N^{2l}} + \int_{|t|>N^{2l}}.$$

As $F_2(t)$ could be bounded by $\ll N^{\frac{1}{2}k}|t|^{\frac{1}{6}}$, recalling the bound (9.12) for I_2 and the bound for I_3 in Proposition 9.9, we bound the first integral as

$$\begin{aligned} \int_{N^{1/10} \leq |t| \leq N^{2l}} &= \int_{N^{1/10} \leq |t| \leq N^{2l}} \min\left(1, \left(\frac{T}{|t|}\right)^{20}\right) |F_1(t)|^2 |F_2(lt)|^2 dt. \\ &\ll \max_{N^{1/10} \leq |t| \leq N^{2l}} \left(\min\left(1, \left(\frac{T}{|t|}\right)\right) |F_2(t)| \right)^2 I_3 \\ &\ll \left(N^{\frac{1}{2}k+\frac{1}{3}l}\right)^2 \left(N^{4l+\epsilon} + \frac{N^{kl+2l+\epsilon}}{\delta} \right) \\ &\ll N^{k+\frac{14}{3}l+\epsilon} + \frac{N^{kl+k+\frac{8}{3}l+\epsilon}}{\delta}. \end{aligned} \tag{9.23}$$

Let $I = [N^{2l}, T]$ or of the form $[T_0, T_0 + T]$, where $T_0 \geq T$. By the high power in $\min(1, (\frac{T}{|t|})^{10})$ it suffices to only consider a single interval I. As in Section 4 of [5], we introduce two level sets

$$\Omega_\lambda = \{|t| \in I; |F_1(t)| \sim \lambda\}$$

and

$$\Omega'_V = \{|t| \in I; |F_2(t)| \sim V\}.$$

Lemma 9.12 implies $|\Omega_\lambda| \ll TN^{2l+\epsilon}\lambda^{-2}$. Comparing this to Lemma 9.13, we find λ can be restricted to $\lambda \leq \lambda_* = N^l T_0^{1/3+\epsilon}$. Letting

$$S(t) = F_2(t)^2 = \sum_{n \sim N^{2l}} a_n n^{it}$$

in Lemma 9.15, where we replace V with V^2, and where $0 \leq |a_n| \ll N^\epsilon$, we find

$$|\Omega'_V| \ll N^\epsilon(N^{4l}V^{-4} + N^{8l}V^{-12}T).$$

We obtain

$$\begin{aligned}
\int_{|t|>N^{2l}} &= \int_{|t|>N^{2l}} \min\left(1, \left(\frac{T}{|t|}\right)^{20}\right) |F_1(t)|^2 |F_2(lt)|^2 dt \\
&\ll \max_{|t|>N^{2l}} \left(\min\left(1, \left(\frac{T}{|t|}\right)\right) |F_2(t)|\right)^2 I_3 \\
&\ll N^\epsilon \max_{\lambda \leq \lambda_*, V} (\lambda^2 V^2)\ |\Omega_\lambda \cap \Omega'_V|.
\end{aligned}$$

Comparing the bounds for $|\Omega_\lambda|$ and $|\Omega'_V|$, we find

$$\lambda^2 V^2 |\Omega_\lambda \cap \Omega'_V| \ll N^\epsilon \min(TN^{2l}V^2, N^{4l}V^{-2}\lambda^2 + TN^{8l}V^{-10}\lambda^2).$$

If the minimum is $TN^{2l}V^2$, then the inequality $TN^{2l}V^2 \leq TN^{8l}V^{-10}\lambda^2$ implies $V^{2l}\lambda^{1/3}$ and thus

$$TN^{2l}V^2 \leq TN^{3l}\lambda^{\frac{1}{3}} \leq TN^{3l}\lambda_*^{\frac{1}{3}}.$$

If the minimum is $N^{4l}V^{-2}\lambda^2 + TN^{8l}V^{-10}\lambda^2$, then the inequality $N^{4l}V^{-2}\lambda^2 \leq TN^{2l}\lambda^{-2}$ implies $V^{-2} \leq T^{\frac{1}{2}}N^{-1}\lambda^{-1}$ and therefore the minimum is bounded by

$$\ll N^{3l}T^{\frac{1}{2}}\lambda \leq N^{3l}T^{\frac{1}{2}}\lambda_*.$$

We find

$$\lambda^2 V^2 |\Omega_\lambda \cap \Omega'_V| \ll N^{3l+\epsilon}T\lambda_*^{\frac{1}{3}} + N^{3l+\epsilon}T^{\frac{1}{2}}\lambda_*.$$

As $T \leq T_0$, we obtain

$$\begin{aligned}
&\ll N^{3l+\epsilon}T_0 N^{\frac{1}{3}l}T_0^{\frac{1}{9}+\epsilon} + N^{3l+\epsilon}T_0^{\frac{1}{2}}N^l T_0^{\frac{1}{3}+\epsilon} \\
&\ll N^{\frac{10}{3}l+\epsilon}T_0^{\frac{10}{9}} + N^{4l+\epsilon}T_0^{\frac{5}{6}}.
\end{aligned}$$

Hence I_2 can be bounded by

$$\begin{aligned} I_2 &\ll N^{-4l-2k}\left(N^{k+\frac{14}{3}l+\epsilon}+\frac{N^{kl+k+\frac{8}{3}l+\epsilon}}{\delta}+N^{\frac{10}{3}l+\epsilon}T^{\frac{10}{9}}+N^{4l+\epsilon}T^{\frac{5}{6}}\right)\\ &\ll N^{-k+\frac{2}{3}l+\epsilon}+N^{kl-k-\frac{4}{3}l+\epsilon}\delta^{-1}+N^{-\frac{2}{3}l-2k+\epsilon}T^{\frac{10}{9}}+N^{-2k+\epsilon}T^{\frac{5}{6}}.\\ &\ll N^{kl-k-\frac{4}{3}l+\epsilon}\delta^{-1}+N^{\frac{10}{9}kl-\frac{2}{3}l-2k+\epsilon}\delta^{-\frac{10}{9}}+N^{\frac{5}{6}kl-2k+\epsilon}\delta^{-\frac{5}{6}} \end{aligned}$$

as $l \geq k \geq 3$ and $\delta > N^{kl-k-2l}$. This proves the proposition. □

As in the proof of Theorem 1.2 *(i)*, we find the smallest δ such that

$$\sum_{N=2^\nu} N^{kl-k-\frac{4}{3}l+\epsilon}\delta^{-1}+N^{\frac{10}{9}kl-\frac{2}{3}l-2k+\epsilon}\delta^{-\frac{10}{9}}+N^{\frac{5}{6}kl-2k+\epsilon}\delta^{-\frac{5}{6}}<\infty.$$

Noting that

$$\begin{aligned} N^{kl-k-\frac{4}{3}l+\epsilon}\delta^{-1}\leq 1 &\iff \delta > N^{kl-k-\frac{4}{3}l}\\ N^{\frac{10}{9}kl-\frac{2}{3}l-2k+\epsilon}\delta^{-\frac{10}{9}}\leq 1 &\iff \delta > N^{kl-\frac{3}{5}l-\frac{9}{5}k}\\ N^{\frac{5}{6}kl-2k+\epsilon}\delta^{-\frac{5}{6}}\leq 1 &\iff \delta > N^{kl-\frac{12}{5}k}. \end{aligned}$$

Using $l \geq k$, we find the smallest δ such that all three inequalities hold is

$$\delta = N^{kl-\frac{12}{5}k+\epsilon}.$$

Using dyadic decomposition and the Borel-Cantelli Lemma 6.9, we obtain Theorem 1.2 *(ii)*.

Remark 9.16. The three lower bounds for δ that we have found above, give us a good view on the importance of the Lindelöf Hypothesis, and on the reason we took the unconditional bound $|\zeta(\frac{1}{2}+it)| \ll (1+|t|)^{\frac{1}{6}}$ for the Riemann zeta function. As we discussed in Section 5, in the last century, there has been many improvements in finding upper bounds for the Riemann zeta function at $\frac{1}{2}+it$. In this proof, we did not choose the smallest bound that has been found so far. However, note that the 'chosen' unconditional bound leads to the first lower bound for δ, i.e.,

$$\delta > N^{kl-k-\frac{4}{3}l}.$$

As we have seen, this is not the smallest bound for δ; the smallest bound is determined by the third inequality and therefore the contribution of the first term is not of main importance. The Lindelöf bound, however, would have contributed to the lower bound

$$\delta > N^{kl-k-2l},$$

which is slightly smaller than the smallest bound for δ in the unconditional case. Therefore, taking the bound for ζ the smallest bound possible, i.e., assuming the Lindelöf Hypothesis, has great influence, but taking a slightly smaller bound for $\zeta(\frac{1}{2}+it)$ than the bound we took, would not contribute to a smaller lower bound for δ. In other words, we could have taken a smaller upper bound for the Riemann zeta function, for example the one used in Section 10, but this would not lead to a different result.

9.2 A polynomial of degree k

We reproduce the steps of the previous section for a more general G_k, namely for

$$G_k(y_1, y_2) := y_1^k + g(y_2),$$

where

$$g(y_2) := \alpha_{2,0} + \alpha_{2,1}y_2 + \ldots + \alpha_{2,k}y_2^k$$

is a polynomial of degree k with $\alpha_{2,k}$ an irrational coefficient. Note this is not a binary form, but it is still continuously differentiable. As y_1^k and $g(y_2)$ are of size N^{kl}, we can apply the results we obtained until now, as all results are based on the size of the terms. Inequality (9.15) can now be expressed as

$$I_4(U) \leq U \sum_{\mathbf{y}\in\mathbb{Z}^4} \tilde{\boldsymbol{\omega}}\left(\frac{\mathbf{y}}{N^l}\right) \mathbf{1}_{|y_1^k+g(y_2)-y_3^k-g(y_4)|\ll\frac{N^{kl}}{U}}.$$

Note that the constant term $\alpha_{2,0}$ cancels out. Again, for $U \ll N^l$, fix y_2, y_3, y_4 in the support of ω_1. As $g(y_2), y_3^k$ and $g(y_4)$ are of order N^{kl}, we find the same bound for $I_4(U)$ for small values of U as in the previous example, which means $I_4(U) \ll N^{4l}$. For large values of U, i.e., for $c_7N^l \leq U \leq T$, with c_7 a sufficiently large positive constant, we apply the results of Huang again [18].

The inequality

$$|y_1^k + g(y_2) - y_3^k - g(y_4)| \ll \frac{N^{kl}}{U}$$

implies

$$y_1^k \in [y_3^k - g(y_2) + g(y_4) - c\frac{N^{kl}}{U}, y_3^k - g(y_2) + g(y_4) + c\frac{N^{kl}}{U}]$$

for some positive constant c. Assuming y_3/N^l lies in the support of ω_1 and $y_2/N^l, y_4/N^l$ lie in the support of ω_2, equations (9.2) imply

$$y_3^k - g(y_2) + g(y_4) \gg N^{kl}.$$

Therefore we can say

$$y_1 \in [(y_3^k - g(y_2) + g(y_4))^{1/k} - c'\frac{N^l}{U}, (y_3^k - g(y_2) + g(y_4))^{1/k} + c'\frac{N^l}{U}] \cap \mathbb{Z}$$

for some positive constant c'. Taking c_7 sufficiently large, this interval can contain at most one integer point. Assume this is the case, then

$$||(y_3^k - g(y_2) + g(y_4))^{1/k}|| \ll \frac{N^l}{U}.$$

We define the function

$$h(z_2, z_4) := (1 - g(z_2) + g(z_4))^{1/k}.$$

We find

$$\begin{aligned}\det(\nabla^2 h) = \; & \frac{1}{k^3}(-g(z_2) + g(z_4) + 1)^{2/k-3} \\ & \times \{(k-1)g''(z_2)g'(z_4)^2 + g''(z_4)(k(g(z_2) - g(z_4) - 1)g''(z_2) - (k-1)g'(z_2)^2)\}.\end{aligned}$$

We choose the domain of h such that h is well-defined and such that $\det \nabla^2 h$ is either strictly positive or negative. Denote this domain with $\mathcal{D}$. We introduce a new weight function $0 \leq \omega_5 \leq 1$ with support in $\mathcal{D}$. Then

$$I_4(U) \leq U \sum_{\substack{(y_2,y_4)\in\mathbb{Z}^4 \\ 1\leq y_3 \leq (b_1 N)^k \\ ||y_4 h(y_2/y_3, y_4/y_3)|| \ll N^l/U}} \omega_5\left(\frac{y_2}{y_3}\right)\omega_5\left(\frac{y_4}{y_3}\right). \tag{9.24}$$

As $\det(\nabla^2 h)$ is either strictly positive or negative in $\mathcal{D}$, we can apply the results of Huang which we explored in Section 8.

Let's compare these results with the previous example. As $g(y_2)$ is also of order N^{kl}, and we have the similar conditions to the weight functions, we obtain exactly the same results. We conclude for large values of U, we have

$$I_4(U) \ll U(\eta N^{3l} + N^{2l+\epsilon})$$

and obtain the following corollary of Theorem 1.2.

Corollary 9.17. *Let $k, l \geq 2$, $k \neq l$. Let $g(x) = \alpha_{2,0} + \alpha_{2,1}x + \ldots + \alpha_{2,k}x^k$ be a polynomial of degree k with $\alpha_{2,k}$ irrational. Then for almost all $\alpha_3 \in [\frac{1}{2}, 1]$, assuming the Lindelöf hypothesis for the Riemann zeta function,*

$$\min_{\substack{x\in\mathbb{Z}^3 \\ |x_1|,|x_2|\sim N^l, |x_3|\sim N^k}} |x_1^k + g(x_2) - \alpha_3 x_3^l| \ll N^{kl-2l-k+\epsilon},$$

for any $\epsilon > 0$, where the constant depends only on $\alpha_{2,k}, \alpha_3$ and ϵ.

Example 9.18. Let $g(y_2) = -\sqrt{2}y_2^3 - 1$, $k = 3$. Then

$$h(z_2, z_4) = (1 + \sqrt{2}z_2^3 - \sqrt{2}z_4^3)^{\frac{1}{3}}$$

and

$$\det(\nabla^2 h) = \frac{-8z_2 z_4}{(\sqrt{2}z_2^3 + \sqrt{2}z_4^3 + 1)^{\frac{7}{3}}}.$$

Putting $|\det(\nabla^2 h)| > 0$ we obtain the solutions

$$\begin{aligned}
z_2 \leq -\frac{1}{\sqrt[6]{2}},\ z_4 &< ((-1)^{2/3}\frac{\sqrt[3]{2z_2^3 + \sqrt{2}}}{\sqrt[3]{2}} \\
-\frac{1}{\sqrt[6]{2}} < z_2 < 0,\ z_4 &< 0 \\
-\frac{1}{\sqrt[6]{2}} < z_2 < 0,\ 0 < z_4 &< ((-1)^{2/3}\frac{\sqrt[3]{2z_2^3 + \sqrt{2}}}{\sqrt[3]{2}} \\
z_2 > 0,\ z_4 &< 0 \\
z_2 > 0,\ 0 < z_4 &< ((-1)^{2/3}\frac{\sqrt[3]{2z_2^3 + \sqrt{2}}}{\sqrt[3]{2}}.
\end{aligned}$$

By choosing $\mathcal{D} = \left[-\frac{1}{\sqrt[6]{2}}, -\epsilon\right]^2$ with $\epsilon > 0$ a small constant, the conditions for Theorem 8.11 are satisfied and the corollary above holds.

9.3 Proof of Theorem 1.3

There are other examples of $G_k(\mathbf{x})$, of which we know how to express y_1 in terms of y_2, y_3, y_4. A well-known example is a quadratic function. Assume k is even and

$$G_k(y_1, y_2) = y_1^k + \alpha_1 y_1^{k/2} y_2^{k/2} + \alpha_2 y_2^k.$$

This relates to understanding the inequality

$$|x_1^k + \alpha_1 x_1^{k/2} x_2^{k/2} + \alpha_2 x_2^k - \alpha_3 x_3^l| < \delta. \tag{9.25}$$

We assume $G_k(y_1, y_2)$ satisfies all conditions we stated in the beginning of the proof. Therefore, we can continue to look for an expression for (9.15), which is now equivalent to finding a bound for

$$I_4(U) \leq U \sum_{\mathbf{y} \in \mathbb{Z}^4} \tilde{\boldsymbol{\omega}} \left(\frac{\mathbf{y}}{N^l} \right) \mathbf{1}_{|y_1^k + \alpha_1 y_2^{k/2} y_1^{k/2} + \alpha_2 y_2^k - y_3^k - \alpha_1 y_3^{k/2} y_4^{k/2} - \alpha_2 y_4^k| \ll \frac{N^{kl}}{U}}.$$

Let $m = k/2$. We can see the inequality as a quadratic function of the variable y_1^m. Let

$$f(y_2, y_3, y_4) := \alpha_2 y_2^{2m} - y_3^{2m} - \alpha_1 y_3^m y_4^m - \alpha_2 y_4^{2m}.$$

Solving

$$-c \frac{N^{2ml}}{U} \leq y_1^{2m} + (\alpha_1 y_2^m) y_1^m + f(y_2, y_3, y_4) \leq c \frac{N^{2ml}}{U}$$

leads to solving two separate quadratic inequalities. We solve the right hand side.

$$\begin{aligned} y_1^{2m} + (\alpha_1 y_2^m) y_1^m + f(y_2, y_3, y_4) &= c \frac{N^{2ml}}{U} \\ y_1^m &= \frac{1}{2} \left(-\alpha_1 y_2^m \pm \sqrt{(\alpha_1 y_2^m)^2 - 4 \left(f(y_2, y_3, y_4) - c \frac{N^{2ml}}{U} \right)} \right). \end{aligned}$$

We assume $D := (\alpha_1 y_2^m)^2 - 4 \left(f(y_2, y_3, y_4) - c \frac{N^{2ml}}{U} \right) > 0$. We consider the solution

$$y_1^m = \frac{1}{2}(-\alpha_1 y_2^m + \sqrt{D}).$$

As D is of order N^{2ml}, we find $\sqrt{D}$ is of size N^{ml}. Let

$$\tilde{f}(y_2, y_3, y_4) := (\alpha_1^2 - 4\alpha_2) y_2^{2m} + 4 y_3^{2m} + 4\alpha_1 y_3^m y_4^m + 4\alpha_2 y_4^{2m}.$$

Then

$$\sqrt{D} = \sqrt{\tilde{f}(y_2, y_3, y_4) + 4c \frac{N^{2ml}}{U}} = \sqrt{N^{2ml}} \sqrt{\frac{\tilde{f}}{N^{2ml}} + \frac{4c}{U}}.$$

Let $d(x) = \sqrt{x}$. Using Taylor expansion, we find

$$\begin{aligned} d\left(\frac{\tilde{f}}{N^{2ml}} + \frac{4c}{U} \right) &= d\left(\frac{\tilde{f}}{N^{2ml}} \right) + \frac{4c}{U} d'\left(\frac{\tilde{f}}{N^{2ml}} \right) + O\left(\left(\frac{4c}{U} \right)^2 d''\left(\frac{\tilde{f}}{N^{2ml}} \right) \right) \\ &= \sqrt{\frac{\tilde{f}}{N^{2ml}}} + \frac{4c}{U} \cdot \frac{1}{2\sqrt{\tilde{f}/N^{2ml}}} + O\left(\left(\frac{4c}{U} \right)^2 d''\left(\frac{\tilde{f}}{N^{2ml}} \right) \right). \end{aligned}$$

For $U \gg N^{ml}$, we find

$$O\left(\left(\frac{4c}{U}\right)^2 d''\left(\frac{\tilde{f}}{N^{2ml}}\right)\right) \leq 3\frac{4c}{U}\frac{\sqrt{N^{2ml}}}{2\sqrt{\tilde{f}}}.$$

Hence

$$\tilde{f}^{1/2} - 4c\tilde{f}^{-1/2} \cdot \frac{N^{2ml}}{U} \leq \sqrt{D} \leq \tilde{f}^{1/2} + 8c\tilde{f}^{-1/2}\frac{N^{2ml}}{U}.$$

We conclude that a solution of $y_1^{2m} + (\alpha_1 y_2^m)y_1^m + f(y_2, y_3, y_4) = c\frac{N^{2ml}}{U}$ satisfies

$$y_1^m \in \left[-\frac{1}{2}\alpha_1 y_2^m + \frac{1}{2}\tilde{f}^{1/2} - 4c\tilde{f}^{-1/2}\frac{N^{2ml}}{U}, -\frac{1}{2}\alpha_1 y_2^m + \frac{1}{2}\tilde{f}^{1/2} + 4c\tilde{f}^{-1/2}\frac{N^{2ml}}{U}\right].$$

Assume $y_2/N^l, y_4/N^l$ lie in the support of ω_2 and y_3/N^l in the support of ω_1. Then $\tilde{f}(y_2, y_3, y_4) \gg N^{2ml}$, so $\tilde{f}^{1/2} \gg N^{ml}$ and $\tilde{f}^{-1/2} \ll N^{-ml}$. Then the solution for y_1 lies in the interval

$$y_1 \in \left[\left(-\frac{1}{2}\alpha_1 y_2^m + \frac{1}{2}\tilde{f}^{1/2}\right)^{\frac{1}{m}} - 4c\frac{N^l}{U}, \left(-\frac{1}{2}\alpha_1 y_2^m + \frac{1}{2}\tilde{f}^{1/2}\right)^{\frac{1}{m}} + 4c\frac{N^l}{U}\right] \cap \mathbb{Z},$$

where c a positive constant. Taking $U \geq c_7 N^l$ sufficiently large, this interval can only contain at most one integer point. Then

$$||(-\tfrac{1}{2}\alpha_1 y_2^m + \tfrac{1}{2}\tilde{f}(y_2, y_3, y_4)^{\frac{1}{2}})^{\frac{1}{m}}|| \ll \tfrac{N^l}{U}.$$

Then the bound for $I_4(U)$ can be rewritten as

$$I_4(U) \leq U \sum_{\substack{(y_2,y_3,y_4)\in\mathbb{Z}^3 \\ (a_2/4)N^l \leq y_2 \leq b_2 N^l \\ ||(-\frac{1}{2}\alpha_1 y_2^m + \frac{1}{2}\tilde{f}(y_2,y_3,y_4)^{\frac{1}{2}})^{\frac{1}{m}}|| \ll \frac{N^l}{U}}} \omega_1\left(\frac{y_3}{N^l}\right)\omega_2\left(\frac{y_4}{N^l}\right).$$

Recall

$$\tilde{f}(y_2, y_3, y_4) = (\alpha_1^2 - 4\alpha_2)y_2^{2m} + 4y_3^{2m} + 4\alpha_1 y_3^m y_4^m + 4\alpha_2 y_4^{2m}.$$

Define

$$\tilde{h}(z_3, z_4) := \tilde{f}(1, y_3/y_2, y_4/y_2)$$

and

$$h(z_3, z_4) := (-\tfrac{1}{2}\alpha_1 + \tfrac{1}{2}\tilde{h}(z_3, z_4)^{\frac{1}{2}})^{\frac{1}{m}}.$$

Let the domain of h be such that h is well-defined and such that the determinant of the Hessian $\nabla^2 h$ is either strictly negative or strictly positive on the domain. Denote this domain with $\mathcal{D}$. Then $h(z_3, z_4)$ has non-vanishing curvature on the domain $\mathcal{D}$ and we can apply the results of Huang [18] again. Furthermore, there exists a smooth weight function $0 \leq \omega_6 \leq 1$ with support in $\mathcal{D}$, such that

$$I_4(U) \leq \sum_{\substack{(y_3,y_4)\in\mathbb{Z}^2 \\ 1 \leq y_2 \leq b_2 N^l \\ ||y_2 h(y_3/y_2, y_4/y_2)|| \ll \frac{N^l}{U}}} \omega_6\left(\frac{y_3}{y_2}\right)\omega_6\left(\frac{y_4}{y_2}\right).$$

Applying Theorem 8.11 with $Q = b_1 N^l$ and $\eta \ll \frac{N^l}{U}$, we obtain

$$I_4(U) \ll U(\eta N^{3l} + N^{2l+\epsilon}).$$

This is the same bound as we found in Section 9.1. Therefore, proceeding as in Section 9.1, we obtain Theorem 1.3.

Example 9.19. Let $G_4(x_1, x_2) = x_1^4 - \frac{1}{\sqrt{2}} x_1^2 x_2^2 + \frac{1}{\sqrt{2}} x_2^4$. With the same notation as in this section, we find

$$f(y_2, y_3, y_4) = \frac{1}{\sqrt{2}} y_2^4 - y_3^4 + \frac{1}{\sqrt{2}} y_3^2 y_4^2 - \frac{1}{\sqrt{2}} y_4^4$$

and

$$\tilde{f}(y_2, y_3, y_4) = (\tfrac{1}{2} - 2\sqrt{2}) y_2^4 + 4y_3^4 - 2\sqrt{2} y_3^2 y_4^2 + 2\sqrt{2} y_4^4.$$

Then

$$\begin{aligned}
\tilde{h}(z_3, z_4) &= \tilde{f}(1, y_3/y_2, y_4/y_2) \\
&= (\tfrac{1}{2} - 2\sqrt{2}) + 4(y_3/y_2)^4 - 2\sqrt{2}(y_3/y_2)^2 (y_4/y_2)^2 + 2\sqrt{2}(y_4/y_2)^4 \\
&= (\tfrac{1}{2} - 2\sqrt{2}) + 4z_3^4 - 2\sqrt{2} z_3^2 z_4^2 + 2\sqrt{2} z_4^4
\end{aligned}$$

and hence

$$h(z_3, z_4) = \left(\frac{1}{2\sqrt{2}} + \frac{1}{2} \left(\left(\frac{1}{2} - 2\sqrt{2} \right) + 4z_3^4 - 2\sqrt{2} z_3^2 z_4^2 + 2\sqrt{2} z_4^4 \right)^{\frac{1}{2}} \right)^{\frac{1}{2}}.$$

We compute the determinant of the Hessian and find

$$\begin{aligned}
\det(\nabla^2 h) =& -\tfrac{1}{4} \Bigg(\frac{2\sqrt{2} z_3 z_4}{\sqrt{-2\sqrt{2} z_3^2 z_4^2 + 2\sqrt{2} z_4^4 + 4z_3^4 - 2\sqrt{2} + \frac{1}{2}} \sqrt{\frac{1}{4}\sqrt{2} + \frac{1}{2}\sqrt{-2\sqrt{2} z_3^2 z_4^2 + 2\sqrt{2} z_4^4 + 4z_3^4 - 2\sqrt{2} + \frac{1}{2}}}} \\
&+ \frac{2\left(\sqrt{2} z_3^2 z_4 - 2\sqrt{2} z_4^3\right)\left(\sqrt{2} z_3 z_4^2 - 4z_3^3\right)}{\left(-2\sqrt{2} z_3^2 z_4^2 + 2\sqrt{2} z_4^4 + 4z_3^4 - 2\sqrt{2} + \frac{1}{2}\right)^{\frac{3}{2}} \sqrt{\frac{1}{4}\sqrt{2} + \frac{1}{2}\sqrt{-2\sqrt{2} z_3^2 z_4^2 + 2\sqrt{2} z_4^4 + 4z_3^4 - 2\sqrt{2} + \frac{1}{2}}}} \\
&- \frac{\left(\sqrt{2} z_3^2 z_4 - 2\sqrt{2} z_4^3\right)\left(\sqrt{2} z_3 z_4^2 - 4z_3^3\right)}{(4\sqrt{2} z_3^2 z_4^2 - 4\sqrt{2} z_4^4 - 8z_3^4 + 4\sqrt{2} - 1)\left(\frac{1}{4}\sqrt{2} + \frac{1}{2}\sqrt{-2\sqrt{2} z_3^2 z_4^2 + 2\sqrt{2} z_4^4 + 4z_3^4 - 2\sqrt{2} + \frac{1}{2}}\right)^{\frac{3}{2}}} \Bigg)^2 \\
&+ \tfrac{1}{4} \Bigg(\frac{2\left(\sqrt{2} z_3^2 z_4 - 2\sqrt{2} z_4^3\right)^2}{\left(-2\sqrt{2} z_3^2 z_4^2 + 2\sqrt{2} z_4^4 + 4z_3^4 - 2\sqrt{2} + \frac{1}{2}\right)^{\frac{3}{2}} \sqrt{\frac{1}{4}\sqrt{2} + \frac{1}{2}\sqrt{-2\sqrt{2} z_3^2 z_4^2 + 2\sqrt{2} z_4^4 + 4z_3^4 - 2\sqrt{2} + \frac{1}{2}}}} \\
&+ \frac{\sqrt{2} z_3^2 - 6\sqrt{2} z_4^2}{\sqrt{-2\sqrt{2} z_3^2 z_4^2 + 2\sqrt{2} z_4^4 + 4z_3^4 - 2\sqrt{2} + \frac{1}{2}} \sqrt{\frac{1}{4}\sqrt{2} + \frac{1}{2}\sqrt{-2\sqrt{2} z_3^2 z_4^2 + 2\sqrt{2} z_4^4 + 4z_3^4 - 2\sqrt{2} + \frac{1}{2}}}} \\
&- \frac{\left(\sqrt{2} z_3^2 z_4 - 2\sqrt{2} z_4^3\right)^2}{(4\sqrt{2} z_3^2 z_4^2 - 4\sqrt{2} z_4^4 - 8z_3^4 + 4\sqrt{2} - 1)\left(\frac{1}{4}\sqrt{2} + \frac{1}{2}\sqrt{-2\sqrt{2} z_3^2 z_4^2 + 2\sqrt{2} z_4^4 + 4z_3^4 - 2\sqrt{2} + \frac{1}{2}}\right)^{\frac{3}{2}}} \Bigg) \\
&\Bigg(\frac{2\left(\sqrt{2} z_3 z_4^2 - 4z_3^3\right)^2}{\left(-2\sqrt{2} z_3^2 z_4^2 + 2\sqrt{2} z_4^4 + 4z_3^4 - 2\sqrt{2} + \frac{1}{2}\right)^{\frac{3}{2}} \sqrt{\frac{1}{4}\sqrt{2} + \frac{1}{2}\sqrt{-2\sqrt{2} z_3^2 z_4^2 + 2\sqrt{2} z_4^4 + 4z_3^4 - 2\sqrt{2} + \frac{1}{2}}}} \\
&+ \frac{\sqrt{2} z_4^2 - 12 z_3^2}{\sqrt{-2\sqrt{2} z_3^2 z_4^2 + 2\sqrt{2} z_4^4 + 4z_3^4 - 2\sqrt{2} + \frac{1}{2}} \sqrt{\frac{1}{4}\sqrt{2} + \frac{1}{2}\sqrt{-2\sqrt{2} z_3^2 z_4^2 + 2\sqrt{2} z_4^4 + 4z_3^4 - 2\sqrt{2} + \frac{1}{2}}}} \\
&- \frac{\left(\sqrt{2} z_3 z_4^2 - 4z_3^3\right)^2}{(4\sqrt{2} z_3^2 z_4^2 - 4\sqrt{2} z_4^4 - 8z_3^4 + 4\sqrt{2} - 1)\left(\frac{1}{4}\sqrt{2} + \frac{1}{2}\sqrt{-2\sqrt{2} z_3^2 z_4^2 + 2\sqrt{2} z_4^4 + 4z_3^4 - 2\sqrt{2} + \frac{1}{2}}\right)^{\frac{3}{2}}} \Bigg)
\end{aligned}$$

This determinant is computed with Sage and shows us an expression for the determinant of the Hessian of $h(z_3, z_4)$ is not straightforward. This motivates why we have not explicitly given an expression for the set $\mathcal{D}$ in the proof of Theorem 1.3.

For simplicity, we choose $\mathcal{D} = [\frac{1}{4}, 1]^2$ to be the domain of (z_3, z_4) in this example. The absolute value of the determinant is strictly positive for $(z_3, z_4) \in \mathcal{D}$ and we can use Huang's results. Therefore, the conditions for Theorem 1.3 are satisfied.

9.4 Conclusion of the three cases

The three examples show us, that the size of the variables x_1, x_2, x_3 is the most important condition on which the smallest solution depends. All these examples give us the same result. Without explicitly proving the general case

$$|G_k(\mathbf{x}) - \alpha_3 x_3^l| < \delta,$$

where $G_k(\mathbf{x}) = x_1^k - \alpha_{2,1} x_1^{k-1} x_2 - \ldots - \alpha_{2,k} x_2^k$, we have a strong suspicion that we will obtain the same results as we obtained for the three examples.

Recall the discussion earlier in this section, regarding the Implicit Function Theorem. This theorem, discussed in Section 7, showed us that it is possible to express y_1 in terms of y_2, y_3, y_4. In order to use Huang's theorem 8.11, we need to assume that the expression $y_1 = \Phi(y_2, y_3, y_4)$ can be written as $y_i \tilde{\Phi}(y_j/y_i, y_k/y_i)$ for $i, j, k \in \{2, 3, 4\}$, $i \neq j \neq k$, with $y_j/y_i, y_k/y_i \in \mathcal{D}$, where $\mathcal{D}$ is a domain on which $\det(\nabla^2 \tilde{\Phi})$ is strictly positive or strictly negative. If this is the case, all necessary conditions hold. We obtain Conjecture 1.4.

In order to prove the conjecture, one has to find a way of performing the same steps as is done in Section 9.1 or 9.2, i.e., to find an upper bound for $I_4(U)$.

10 Proof of Theorem 1.5

As discussed in Section 6.2, one way of exploring the distribution of the values of polynomials in a Diophantine inequality, is by looking at the smallest gap between two values. In the article of Blomer, Bourgain, Radziwiłł and Rudnick [2], this is done for the Diophantine inequality $\alpha m^2 + n^2 \leq X$. Let α be an irrational real number. Let λ_j be the eigenvalues of the Dirichlet Laplacian of a rectangular billiard with width $\pi/\sqrt{\alpha}$ and height π. Then each λ_j is of the form $\alpha m^2 + n^2$ with integers $m, n \geq 1$. Let

$$\#\{j : \lambda_j \leq X\} = \#\{(m,n) : m, n \geq 1, \alpha m^2 + n^2 \leq X\}.$$

We study the size of the minimal gap between two λ_i's. Therefore, we define

$$\delta_{\min}^{(\alpha)}(N) = \min(\lambda_{i+1} - \lambda_i : 1 \leq i < N).$$

If $\alpha m^2 + n^2 \leq X$, then both m, n are of size $\ll X^{\frac{1}{2}}$. This means there are at most $\ll X^{\frac{1}{2}}$ different values for m and n. This, in its turn, implies that there are also $\ll X$ values for $\lambda_j \leq X$. In particular, $\lambda_X \ll X$. Therefore, we expect the values λ_i, λ_{i+1} in $\delta_{\min}^{(\alpha)}(N)$ to be of size $\sqrt{N}$. In conclusion, by finding a minimum for

$$|(\alpha m^2 + n^2) - (\alpha m'^2 + n'^2)|,$$

where $m, m', n, n' \ll \sqrt{N}$, we at least find an upper bound for $\delta_{\min}^{(\alpha)}(N)$.

In [2, Proposition 2.2], some results on the lower bound of $\delta_{\min}^{(\alpha)}(N)$ are given. Blomer, Bourgain, Radziwiłł and Rudnick explain that the hardest part is to give an upper bound. A simple argument shows we can at least expect an upper bound

$$\delta_{\min}^{(\alpha)}(N) \ll N^{-\frac{1}{2}}$$

for any irrational $\alpha > 0$. Let $Q \geq 1$ be sufficiently large. Recall Dirichlet's approximation theorem: there exist $a \in \mathbb{Z}$, $1 \leq q \leq Q$ such that $0 < |a - q\alpha| \leq 1/Q$. As $\alpha > 0$, we find $a \geq 1$. Define $m = 2q+1, m' = 2q-1$, and $n = 2a-1, n' = 2a+1$. Then $1 \leq m, m', n, n' \ll Q$ and

$$|\alpha m^2 + n^2 - (\alpha m'^2 + n'^2)| = |8\alpha q - 8a| = 8|\alpha q - a| \leq \frac{8}{Q}.$$

As $m, m', n, n' \ll N^{1/2}$, we choose $Q = N^{1/2}$ and obtain the result. In [2], a better result for an upper bound is obtained and is given in Theorem 1.5. In this section, we prove this theorem.

We first prove the following proposition.

Proposition 10.1. *Let $\mathcal{J} \subset (0, \infty)$, $\alpha \in \mathcal{J}$, $M \geq 1$ real and $M^3 \leq T \leq M^4$. Define*

$$S(M, T, \alpha) := \#\left\{n_1, n_2, n_3, n_4 \asymp M : \left|\frac{n_1 n_2}{n_3 n_4} - \alpha\right| \ll \frac{1}{T}\right\}.$$

Then for any $\eta > 0$ sufficiently small, we have $S(M, M^{4-\eta}, \alpha) \geq 1$ for all sufficiently large $M \geq M_0(\alpha)$, and all $\alpha \in \mathcal{J} \backslash \mathcal{T}_M$, where $\mathcal{T}_M$ is an exceptional set of measure $\mu(T_M) \ll M^{-\rho}$ with $\rho > 0$ depending only on $\eta > 0$.

This quantity $S(M,T,\alpha)$ is related to $\delta_{\min}^{(\alpha)}(N)$ in the following way. By writing $n_1 = m - m'$, $n_2 = m + m'$ and $n_3 = n - n'$, $n_4 = n + n'$, we find $n_1 n_2 = m^2 - m'^2$ and $n_3 n_4 = n^2 - n'^2$, so by estimating $S(M,T,\alpha)$, we obtain the number of m, m', n, n' that satisfy the inequality for n_1, n_2, n_3, n_4. As we want this number to be minimal, we seek to find a lower bound for $S(M,T,\alpha)$ for almost all α and T as large as possible in terms of M.

Remark 10.2. As this thesis is about generalising Diophantine inequalities, one could ask what will happen for the inequality

$$\alpha m^k + n^2 \leq X,$$

where $k \geq 2$. In this case, we want to minimise

$$|(\alpha m^k + n^2) - (\alpha m'^k + n'^2)|.$$

As in the quadratic case, we write

$$n^2 - n'^2 = (n - n')(n + n') =: n_2 n_3.$$

Furthermore, we can write the k-th powers as

$$m^k - m'^k = (m - m')(m^{k-1} + m^{k-2}m' + \ldots + mm'^{k-2} + m'^{k-1}).$$

Denote $n_1 := m - m'$ and $p_{k-1} := m^{k-1} + m^{k-2}m' + \ldots + mm'^{k-2} + m'^{k-1}$. In this case, the counting function as defined in Proposition 10.1, will be of the form

$$S(M,T,\alpha) := \#\left\{ n_1, n_2, n_3 \asymp M, p_{k-1} \asymp M^{k-1} : \left| \frac{n_1 p_{k-1}}{n_2 n_3} - \alpha \right| \ll \frac{1}{T} \right\}.$$

Note p_{k-1} is now a polynomial dependent of n_1 and another variable which we denote with n'.

We first prove Proposition 10.1. After that, we show how Theorem 1.5 can be derived using this proposition.

Proof of Proposition 10.1. Similar to the proof of Theorem 1.2, we introduce two non-negative smooth weight functions ω_7, ω_8 that are bounded by 1. Let $\omega_7 = 1$ on $[a, b]$ with $0 < a < b$ and $\operatorname{supp}(\omega_7) \subset [\frac{1}{2}a, 2b]$. Let $\omega_8 = 1$ on $[-1, 1]$ and $\operatorname{supp}(\omega_8) \subset [-2, 2]$. Fix some small $\eta > 0$ and let $\epsilon > 0$ denote an arbitrarily small constant, not necessarily the same each time it occurs. Define

$$\tilde{S}(M,T,\alpha) := \#\left\{ n_i \asymp M^{\eta/4}, m_i \asymp M^{1-\eta/4} : \left| \frac{n_1 m_1 n_2 m_2}{n_3 m_3 n_4 m_4} - \alpha \right| \ll \frac{1}{T} \right\}.$$

By the standard divisor bound, we obtain

$$S(M,T,\alpha) \gg M^{-\epsilon} \tilde{S}(M,T,\alpha).$$

We are going to bound $\tilde{S}$ from below. Let $\beta = \log \alpha$. Note that $\tilde{S}$ can be counted by

$$\sum_{\substack{n_1,n_2,n_3,n_4 \\ m_1,m_2,m_3,m_4}} \prod_{i=1}^{4} \omega_7\left(\frac{n_i}{M^{\eta/4}}\right) \omega_7\left(\frac{m_i}{M^{1-\eta/4}}\right) \mathbf{1}_{\left|\frac{n_1 m_1 n_2 m_2}{n_3 m_3 n_4 m_4} - \alpha\right| \ll \frac{1}{T}}.$$

Let us compare this strategy to the proof of Theorem 1.2. Recall the definitions of $F_1(t)$ and $F_2(t)$ in equations (9.5) and (9.6) respectively. Let $\widehat{\omega_8}(y)$ be the Fourier transform of ω_8 again. We define these expressions now as

$$F_1(t) = \sum_{\substack{n_1,n_2,n_3,n_4 \\ m_1,m_2,m_3,m_4}} \prod_{i=1}^{4} \omega_7\left(\frac{n_i}{M^{\eta/4}}\right)\omega_7\left(\frac{m_i}{M^{1-\eta/4}}\right) e^{it\log\left(\frac{n_1m_1n_2m_2}{n_3m_3n_4m_4}\right)} \tag{10.1}$$

$$F_2(t) = e^{it\log(\alpha)}. \tag{10.2}$$

Repeating the steps of the proof of Theorem 1.2, we obtain

$$\begin{aligned}
&\tfrac{1}{T}\textstyle\int_{\mathbb{R}} \widehat{\omega_8}\left(\tfrac{t}{T}\right) F_1(t)\overline{F_2(t)}dt \\
&\quad = \tfrac{1}{T}\textstyle\sum\prod \omega_7\left(\tfrac{n_i}{M^{\eta/4}}\right)\omega_7\left(\tfrac{m_i}{M^{1-\eta/4}}\right)\int_{\mathbb{R}}\widehat{\omega_8}\left(\tfrac{t}{T}\right) e^{it\left(\log\left(\frac{n_1m_1n_2m_2}{n_3m_3n_4m_4}\right)-\log(\alpha)\right)} \\
&\quad \geq \textstyle\sum\prod \omega_7\left(\tfrac{n_i}{M^{\eta/4}}\right)\omega_7\left(\tfrac{m_i}{M^{1-\eta/4}}\right)\omega_8\left(T\left(\log\left(\frac{n_1m_1n_2m_2}{n_3m_3n_4m_4}\right)-\log(\alpha)\right)\right).
\end{aligned}$$

Hence, a lower bound of $\tilde{S}(M,T,\alpha)$ is given by

$$\tilde{S}(M,T,\alpha) \geq \textstyle\sum\omega_8\left(T\left(\log\frac{n_1m_1n_2m_2}{n_3m_4n_4m_4} - \log\alpha\right)\right)\prod\omega_7\left(\frac{n_i}{M^{\eta/4}}\right)\omega_7\left(\frac{m_i}{M^{1-\eta/4}}\right).$$

In [2], the lower bound is written as

$$\tfrac{1}{T}\textstyle\int_{-\infty}^{\infty}\widehat{\omega_8}\left(\tfrac{y}{T}\right)e^{-2\pi iy\beta}\left|\sum_n \omega_7\left(\tfrac{n}{M^{\eta/4}}\right)n^{2\pi iy}\right|^4\left|\sum_m\omega_7\left(\tfrac{m}{M^{1-\eta/4}}\right)m^{2\pi iy}\right|^4 dy.$$

This equation is explained by

$$\begin{aligned}
&\omega_8\left(T\left(\log\frac{n_1m_1n_2m_2}{n_3m_3n_4m_4}-\beta\right)\right) \\
&= \int_{-\infty}^{\infty}\widehat{\omega_8}\left(\frac{y}{T}\right)e^{2\pi i(\log\frac{n_1m_1n_2m_2}{n_3m_3n_4m_4}-\beta)y}\frac{dy}{T} \\
&= \frac{1}{T}\int\widehat{\omega_8}\left(\frac{y}{T}\right)e^{-2\pi i\beta y}e^{2\pi i\log(n_1m_1n_2m_2)}e^{-2\pi i\log(n_3m_3n_4m_4)}dy \\
&= \frac{1}{T}\int\widehat{\omega_8}\left(\frac{y}{T}\right)e^{-2\pi i\beta y}(n_1m_1n_2m_2)^{2\pi iy}(n_3m_3n_4m_4)^{-2\pi iy}dy.
\end{aligned}$$

Multiplying by

$$\prod_{i=1}^{4}\omega_7\left(\frac{n_i}{M^{\eta/4}}\right)\omega_7\left(\frac{m_i}{M^{1-\eta/4}}\right)$$

and summing over $n_1,\dots,n_4$, $m_1,\dots,m_4$ gives us the right expression.

We split the integral in two parts $I_1(\beta)+I_2(\beta)$, where $I_1(\beta)$ is the integral for $|y|\leq M^\epsilon$ for some very small fixed $\epsilon>0$. $I_2(\beta)$ is the other part of the integral. Again, we will show $I_1(\beta)$ is the main part and $I_2(\beta)$ can be considered as the error term. We let $\check{\omega}_7(s)$ be the Mellin transform of ω_7 (see Section 5). Define

$$\Sigma(N,y) := \sum_n \omega_7\left(\frac{n}{N}\right)n^{2\pi iy} = \int_{2-i\infty}^{2+i\infty}\check{\omega}_7(s)N^s\zeta(s-2\pi iy)\frac{ds}{2\pi i}, \tag{10.3}$$

where $\zeta(s)$ is the Riemann zeta function. As argued before, $I_1(\beta)$ has rapid decay along the vertical lines, so for $|y| \le M^\epsilon$, $N = M^c$ for $c = \eta/4$ or $c = 1 - \eta/4$, we find

$$\Sigma(N, y) = \breve{\omega}_7(1 + 2\pi i y)N^{1+2\pi i y} + O(N^\epsilon). \tag{10.4}$$

We obtain

$$\begin{aligned} I_1(\beta) &= \frac{1}{T}\int_{-M^\epsilon}^{M^\epsilon} \widehat{\omega}_8\left(\tfrac{y}{T}\right) e^{-2\pi i y\beta} \left|\textstyle\sum_n \omega_7\left(\frac{n}{M^{\eta/4}}\right) n^{2\pi i y}\right|^4 \left|\textstyle\sum_m \omega_7\left(\frac{m}{M^{1-\eta/4}}\right) m^{2\pi i y}\right|^4 dy \\ &= \frac{1}{T}\int_{-M^\epsilon}^{M^\epsilon} \widehat{\omega}_8\left(\frac{y}{T}\right) e^{-2\pi i y\beta} \left|\Sigma(M^{\eta/4}, y)\right|^4 \left|\Sigma(M^{1-\eta/4}, y)\right|^4 \\ &= \frac{M^4}{T}\int_{-M^\epsilon}^{M^\epsilon} \widehat{\omega}_8\left(\frac{y}{T}\right) e^{-2\pi i y\beta} |\breve{\omega}_7(1+2\pi i y)|^8 dy + \frac{1}{T}O(N^{4-\eta/4+\epsilon}). \end{aligned}$$

Define

$$c(\beta) = \widehat{\omega}_8(0)\int_{-\infty}^{\infty} e^{-2\pi i y\beta}|\breve{\omega}_7(1+2\pi i y)|^8 dy.$$

By Taylor's theorem, we obtain

$$I_1(\beta) = c(\beta)\frac{M^4}{T} + O\left(\frac{M^{4-4/\eta+\epsilon}}{T} + \frac{M^4}{T^2}\right).$$

As in [2], we define $v(t) := w_7(e^t)e^t$, which is a non-negative compactly supported function. Then

$$\begin{aligned} \hat{v}(-t) &= \int_{-\infty}^{\infty} v(y)e^{2\pi i t y} dy \\ &= \int_{-\infty}^{\infty} \omega_7(e^y)e^y e^{2\pi i t y} dy \\ &= \int_{-\infty}^{\infty} \omega_7(e^y)e^{y(1+2\pi i t)} dy. \end{aligned}$$

Substitute $x = e^y$, $dy = \frac{1}{x}dx$. Then

$$\begin{aligned} \hat{v}(-t) &= \int_{-\infty}^{\infty} \omega_7(x)e^{(\log x)(1+2\pi i t)}\frac{1}{x}dx \\ &= \int_{-\infty}^{\infty} \omega_7(x)x^{2\pi i t} dx \\ &= \breve{\omega}_7(1+2\pi i t). \end{aligned}$$

We obtain

$$c(\beta) = \widehat{\omega_8}(0)\int_{\mathbb{R}^7} v(t_1)\ldots v(t_7)v(-\beta + t_1 + \ldots + t_4 - t_5 - t_6 - t_7)dt_1\cdots dt_7.$$

Choosing the support of ω_7 sufficiently large, we make sure $c(\beta)$ is bounded away from 0, uniformly for all e^β in the domain $\mathcal{J}$. Then

$$I_1(\beta) \gg \frac{M^4}{T},$$

uniformly in β.

The last part of the proof is showing that the contribution of $I_2(\beta)$, i.e., the contribution of large frequencies $|y| > M^\epsilon$, is of lower order of magnitude, for almost all β. Define

$$\mathcal{I} := \left(\int_{\log \mathcal{J}} |I_2(\beta)|^2 d\beta \right)^{\frac{1}{2}}. \tag{10.5}$$

Also, define

$$F(y) := \mathbf{1}_{|y|>M^\epsilon} \frac{1}{T} \widehat{\omega_8} \left(\frac{y}{T} \right) \left| \Sigma(M^{\eta/4}, y) \Sigma(M^{1-\eta/4}, y) \right|^4.$$

Then

$$\begin{aligned} I_2(\beta) &= \int_{|y|>M^\epsilon} \widehat{\omega_8} \left(\frac{y}{T} \right) e^{-2\pi i y \beta} \left| \Sigma(M^{\eta/4}, y) \Sigma(M^{1-\eta/4}, y) \right|^4 dy. \\ &= \int_{-\infty}^{\infty} F(y) e^{-2\pi i \beta y} dy \\ &= \hat{F}(\beta). \end{aligned}$$

By Parseval's theorem 3.6, we find

$$\mathcal{I}^2 \le \int_{-\infty}^{\infty} |I_2(\beta)|^2 d\beta = \int_{-\infty}^{\infty} |F(y)|^2 dy.$$

Therefore,

$$\mathcal{I}^2 \ll \frac{1}{T^2} \int_{|y|>M^\epsilon} \left| \widehat{\omega_8} \left(\frac{y}{T} \right) \right|^2 \left| \Sigma(M^{\eta/4}, y) \right|^8 \left| \Sigma(M^{1-\eta/4}, y) \right|^8 dy. \tag{10.6}$$

We can bound $\Sigma(M^{\frac{\eta}{4}}, y)$ in a similar way as we did in Section 9.1. We shift the contour in (10.3) to $\mathrm{Re}\, s = 1 - \eta^4$. The pole at $s = 1 + 2\pi i y$ falls in the contour. As $|y| \ge M^\epsilon$, we find $\check{\omega}_7$ has rapid decay along the vertical lines. Therefore, the contribution of this pole to the integral is negligible. Using the same strategy as in Section 9.1, we find the following upper bound.

$$\Sigma(M^{\eta/4}, y) \ll M^{\eta/4(1-\eta^4)} \int_{-\infty}^{\infty} \frac{\zeta(1 - \eta^4 - 2\pi i y + it)}{1 + |t|^{10}} dt.$$

We use the bound

$$|\zeta(\sigma + it)| \ll |t|^{A(1-\sigma)^{3/2}+\epsilon},$$

with $\frac{1}{2} \le \sigma \le 1$, $|t| \ge 2$ as opposed in [2]. This is valid as $y \ge \frac{1}{2\pi} t$ and $y \ge M^\epsilon$, hence $t \ge M^\epsilon$, so by taking M sufficiently large, $|t| \ge 2$.As $\sigma = 1 - \eta^4$, we obtain

$$|\Sigma(M^{\eta/4}, y)| \ll M^{\eta/4(1-\eta^4)} |y|^{A\eta^6+\epsilon}.$$

And therefore

$$\begin{aligned} \mathcal{I}^2 &\ll \frac{1}{T^2} M^{2\eta(1-\eta^4)} \int_{\mathbb{R}} \left| \widehat{\omega_8} \left(\frac{y}{T} \right) \right|^2 |y|^{8A\eta^6+\epsilon} |\Sigma(M^{1-\eta/4}, y)|^8 dy \\ &\ll \frac{1}{T^2} M^{2\eta(1-\eta^4)} T^{8A\eta^6+\epsilon} \int_{|y| \le T^{1+\epsilon}} \left| \sum_{n \ll M^{4-\eta}} a(n) n^{2\pi i y} \right|^2 dy. \end{aligned}$$

Here

$$a(n) = \sum_{n_1 \cdot \ldots \cdot n_4 = n} \omega_7\left(\frac{n_1}{M^{1-\eta/4}}\right) \cdots \omega_7\left(\frac{n_4}{M^{1-\eta/4}}\right) \ll n^{\epsilon}.$$

We use the standard mean value theorem as stated in Theorem 9.1 in [19]. This is

$$\int_0^X \left| \sum_{n \le N} a_n n^{it} \right|^2 dt \ll (X + N) \sum_{n \le N} |a_n|^2.$$

With $N = M^{4-\eta}$, $X = T^{1+\epsilon}$ and $a_n = a(n)$, we obtain

$$\begin{aligned}
\mathcal{I}^2 &\ll M^{2\eta(1-\eta^4)} \frac{1}{T^2} T^{8A\eta^6+\epsilon} (T^{1+\epsilon} + M^{4-\eta}) \sum_{n \ll M^{4-\eta}} |n^{\epsilon}|^2 \\
&\ll M^{2\eta(1-\eta^4)} \frac{1}{T^2} (M^{4-\eta})^{8A\eta^6+\epsilon} (M^{4-\eta+\epsilon} + M^{4-\eta}) \cdot M^{4-\eta} \\
&\ll M^{2\eta(1-\eta^4)} \frac{1}{T^2} (M^{4-\eta})^{8A\eta^6+\epsilon} M^{8-2\eta} \\
&\ll \frac{1}{T^2} M^{-2\eta^5+32A\eta^6+8}.
\end{aligned}$$

For $\eta > 0$ sufficiently small, $\eta^5 > 32A\eta^6 + \epsilon$. We obtain

$$\mathcal{I}^2 \ll \frac{1}{T^2} M^{8-\eta^5}.$$

Set $\rho = \frac{1}{2}\eta^5$. Then

$$\mathcal{I} \ll M^{4-\rho} T^{-1}, \tag{10.7}$$

where $T = M^{4-\eta}$. We claim this implies that

$$I_2(\beta) \ll M^{4-\rho/2} T^{-1}$$

for all β except for a small set $\mathcal{T}_M$ of measure $\ll M^{-\rho}$. Let

$$B := M^{4-\rho/2} T^{-1} = M^{\eta-\rho/2}.$$

As

$$\int_{|I_2(\beta)|>B} |I_2(\beta)|^2 d\beta \ge \int_{|I_2(\beta)|>B} |I_2(\beta)|^2 d\beta \ge B^2 \cdot \mu\{\beta : I_2(\beta) > B\},$$

we obtain

$$\mu\{\beta : I_2(\beta) > B\} = \int_{|I_2(\beta)|>B} 1 \, d\beta \ll \mathcal{I}^2.$$

This results in the following inequality

$$\begin{aligned}
\int_{|I_2(\beta)|>B} 1 \, d\beta &\le \int \frac{|I_2(\beta)|^2}{B^2} d\beta \\
&= \frac{\mathcal{I}^2}{B^2} \\
&\ll \frac{M^{2\eta-2\rho}}{M^{2\eta-\rho}} = M^{-\rho}.
\end{aligned}$$

The claim is hereby proved. Hence for all $\alpha \in \mathcal{J} \backslash \mathcal{T}_M$,

$$\begin{aligned} S(M, M^{4-\eta}, \alpha) &\gg M^{-\epsilon} \tilde{S}(M, M^{4-\eta}, \alpha) \\ &\gg M^{-\epsilon} \left(\frac{M^4}{T} + O\left(\frac{M^{4-\rho}}{T} \right) \right) \\ &\gg \frac{M^{4-\epsilon}}{T} \geq 1. \end{aligned}$$

This completes the proof. $\square$

Remark 10.3. Generalising the proof for the case $\alpha m^k + n^2$ is not so straightforward. Recall the discussion in Remark 10.2. One could think of defining $\tilde{S}(M, T, \alpha)$ as

$$\tilde{S}(M, T, \alpha) = \# \left\{ n_i \asymp M^{\eta/3}, m_i \asymp M^{1-\eta/3}, p_{k-1} \asymp M^{k-1} : \left| \tfrac{n_1 m_1 p_{k-1}}{n_2 m_2 n_3 m_3} - \alpha \right| \ll \tfrac{1}{T} \right\}.$$

The expressions $F_1(t)$ and $F_2(t)$ as in (10.1) and (10.2) can be defined as

$$\begin{aligned} F_1(t) &= \sum_{\substack{n_1, n_2, n_3, \\ m_1, m_2, m_3}} \prod_{i=1}^{3} \omega_7 \left(\frac{n_i}{M^{\eta/3}} \right) \omega_7 \left(\frac{m_i}{M^{1-\eta/3}} \right) e^{it \log\left(\frac{n_1 m_1 p_{k-1}}{n_2 m_2 n_3 m_3} \right)} \\ F_2(t) &= e^{it \log(\alpha)}. \end{aligned}$$

The interested reader is invited to think of ways how to proceed in solving this problem.

For the proof of Theorem 1.5, we need one more lemma. Here an integer n is called evenly divisible if there exists a divisor d of n such that $\min(d, n/d) \gg n^{1/2-\epsilon}$ for all $\epsilon > 0$.

Lemma 10.4 (Lemma 3.2 in [2]). *Assume*

$$\left| \alpha - \frac{p}{q} \right| \ll \frac{1}{T}$$

holds for some $T \leq q^2$. If $\alpha > 0$ has infinitely many good rational approximations p_n/q_n with $q_1 < q_2 < \ldots$, for this inequality, with both p and q evenly divisible and $q_n \geq c q_{n+1}$ for some constant $c > 0$ and all $n \geq 1$, then

$$\delta_{\min}^{(\alpha)}(N) \ll N^{1+\epsilon} T^{-1}$$

for all N and all $\epsilon > 0$.

Using this lemma and Proposition 10.1, we are able to prove Theorem 1.5.

Proof of Theorem 1.5. The proof of the theorem is very similar to other techniques we used in Section 9: we use dyadic decomposition and the Borel-Cantelli Lemma 6.9 again. Assume Proposition 10.1 holds. Let $M = 2^\nu$, $\nu \in \mathbb{N}$. Then

$$\sum_{M=2^\nu} \mu(\mathcal{T}_M) \ll \sum_{M=2^\nu} M^{-\rho} = \sum_{M=2^\nu} 2^{-\rho\nu} < \infty.$$

We now use the Borel-Cantelli Lemma 6.9 and conclude $S(M, M^{4-\eta}, \alpha) \geq 1$ for almost all α, all sufficiently large $M = 2^\nu \geq M_0(\alpha)$ and $\eta > 0$ sufficiently small. Hence by the Borel-Cantelli Lemma, we have an exceptional set independent of M. By a similar argument as in the proofs of Section 9, using dyadic decomposition, the Borel-Cantelli Lemma holds for all M sufficiently large. Since Proposition 10.1 holds, we can apply Lemma 10.4. As we took $n_1, \ldots n_4 \asymp M$ in our definition for $S(M, T, \alpha)$, we find $T = M^{4-\eta} = (N^{\frac{1}{2}})^{4-\eta}$. It follows

$$\delta_{\min}^{(\alpha)}(N) \ll N^{1+\epsilon} N^{\frac{\eta-4}{2}} = N^{-1+\eta/2+\epsilon}. \tag{10.8}$$

for all sufficiently large integers $N \geq N_0(\alpha)$. As the implied constant is allowed to depend on α, this equation holds in fact for all N. Then for η sufficiently small, we obtain the bound

$$\delta_{\min}^{(\alpha)}(N) \ll \frac{1}{N^{1-\epsilon}}$$

for all $\epsilon > 0$ and all N. This completes the proof. □

11 Connection to Hardy-Littlewood circle method

An often used tool in analytic number theory, for studying rational points on higher-dimensional algebraic varieties, is the Hardy-Littlewood circle method. Let $f \in \mathbb{Z}[x_1, \ldots, x_n]$ be a homogeneous polynomial of degree d. The method concentrates on the counting function

$$N(f; B) := \#\{\mathbf{x} \in \mathbb{Z}^n : f(\mathbf{x}) = 0, |\mathbf{x}| \leq B\}$$

for any $B \geq 1$. Important here is that the number of variables is sufficiently large in terms of the degree. As this looks very similar to the counting problems of this thesis, we will shortly introduce the concepts of the Hardy-Littlewood circle method and explore the similarities between the two methods. Since our original problem is not counting rational points *on* an algebraic variety, but we are counting rational points *close to* it, we are not able to use the Hardy-Littlewood circle method in Section 9.

11.1 Outline of the circle method

Without exploiting the technical details of the Hardy-Littlewood circle method, we will outline the significant steps that are needed. Two well-readable references which provide more technicalities are Chapter 8 of Browning [6] and *Analytic methods for Diophantine equations and Diophantine inequalities* from Davenport [9]. Given B, the starting point of the Hardy-Littlewood circle method is the generating function

$$S(\alpha) = \sum_{\mathbf{x} \in \mathbb{Z}^n \cap [-B,B]^n} e(\alpha f(\mathbf{x})).$$

Here $\alpha \in [0, 1]$ is a real variable. The strength of this generating function is the identity

$$\int_0^1 e(\alpha n) d\alpha = \begin{cases} 1, & \text{if } n = 0 \\ 0, & \text{if } n \in \mathbb{Z}\backslash\{0\}. \end{cases} \tag{11.1}$$

This suggests there will often be significant cancellation. Note

$$\begin{aligned} \int_0^1 S(\alpha) d\alpha &= \int_0^1 \sum_{\mathbf{x} \in \mathbb{Z}^n \cap [-B,B]^n} e(\alpha f(\mathbf{x})) d\alpha \\ &= \sum_{\mathbf{x} \in \mathbb{Z}^n \cap [-B,B]^n} \int_0^1 e(\alpha f(\mathbf{x})) d\alpha \\ &= \#\{\mathbf{x} \in \mathbb{Z}^n : f(\mathbf{x}) = 0, |\mathbf{x}| \leq B\}, \end{aligned}$$

thus we obtain the expression

$$N(f; B) = \int_0^1 S(\alpha) d\alpha. \tag{11.2}$$

The next step of the Hardy-Littlewood circle method is motivated by the observation that $S(\alpha)$ can be rather large for values of $\alpha \in (0, 1)$ that are well-approximated by a rational number $\frac{a}{q}$ with small denominator and one expects

$S(\alpha)$ to be small for such α that are not well-approximated by rationals with small denominator. To motivate the first observation, we look at $S(0) = S(\frac{0}{1})$. Note that

$$\begin{aligned} S(0) &= \sum_{x \in \mathbb{Z}^n \cap [-B,B]^n} e(0) \\ &= \#\{\mathbf{x} \in \mathbb{Z}^n : |\mathbf{x}| \leq B\} = o(B^n) \end{aligned}$$

and this is obviously the largest $S(\alpha)$ we could get. One can show that for any fraction $\frac{a}{q} \in \mathbb{Q}$, not just the value 0, the sum $S(\frac{a}{q})$ will be of exact order B^n. To see this, we introduce

$$S_{q,a} := \sum_{\mathbf{x} \mod q} e_q(af(\mathbf{x})),$$

where $e_q(a) := e(\frac{a}{q})$. As $e_q(af(\mathbf{x})) = e_q(af(\mathbf{y}))$ whenever $\mathbf{x} \equiv \mathbf{y} \mod q$, we find

$$S\left(\frac{a}{q}\right) = \sum_{\mathbf{x} \mod q} e_q(af(\mathbf{x}))\#\{\mathbf{y} \in \mathbb{Z}^n : |\mathbf{y}| \leq B, \ \mathbf{y} \equiv \mathbf{x} \mod q\}.$$

Furthermore, since

$$\#\{\mathbf{y} \in \mathbb{Z}^n : |\mathbf{y}| \leq B, \ \mathbf{y} \equiv \mathbf{x} \mod q\} = \left(\frac{2B}{q} + O(1)\right)^n = \frac{2^n B^n}{q^n} + O(B^{n-1}),$$

we obtain

$$S\left(\frac{a}{q}\right) = \frac{2^n B^n}{q^n} + O(q^n B^{n-1}).$$

This shows why $S(\frac{a}{q})$ is of exact order B^n if $S_{a,q} \neq 0$. Extending this analysis, one can show that $S(\alpha)$ is also of exact order B^n when α is close to such $\frac{a}{q}$. Therefore, write $\alpha = \frac{a}{q} + z$. Similar to what we did above, write

$$S\left(\frac{a}{q} + z\right) = \sum_{\mathbf{x} \mod q} e_q(af(\mathbf{x})) \sum_{\substack{\mathbf{y} \in \mathbb{Z}^n \cap [-B,B]^n \\ \mathbf{y} \equiv \mathbf{x} \mod q}} e(zf(\mathbf{y})).$$

The following lemma is proved in [6] and gives us insightful results about this sum.

Lemma 11.1 (Lemma 8.2 in [6])**.** *Let $a, q \in \mathbb{Z}$ such that $1 \leq a \leq q \leq B$ and* $\gcd(a, q) = 1$. *We have*

$$S\left(\frac{a}{q} + z\right) = q^{-n} B^n S_{q,a} I(zB^d) + O(qB^{n-1}(1 + |z|B^d)),$$

where $I(\gamma) := \int_{\mathbf{t} \in [-1,1]^n} e(\gamma f(\mathbf{t})) d\mathbf{t}$.

It follows that for fixed $\frac{a}{q}$ such that $S_{q,a} \neq 0$, the exponential sum $S(\alpha)$ has exact order B^n when z is sufficiently small as we write $\alpha = \frac{a}{q} + z$.

The second observation, about the smaller order of badly approximable α, comes from Weyl's inequality and Hua's lemma, which are stated below as they are given and proved in [27].

Theorem 11.2 (Weyl's inequality)**.** *Let $f(x) = \alpha x^k + \ldots$ be a polynomial of degree $k \geq 2$ with real coefficients and suppose α has rational approximation $\frac{a}{q}$ such that $|\alpha - \frac{a}{q}| \leq \frac{1}{q^2}$, with $q \geq 1$, $\gcd(a, q) = 1$. Let $S(f) = \sum_{n=1}^{N} e(f(n))$. Let $K = 2^{k-1}$ and $\epsilon > 0$. Then*

$$S(f) \ll N^{1+\epsilon}(N^{-1} + q^{-1} + N^{-1}q)^{1/K}.$$

Theorem 11.3 (Hua's lemma)**.** *Let $k \geq 2$, and $T(\alpha) = \sum_{n=1}^{N} e(\alpha n^k)$. Then*

$$\int_0^1 |T(\alpha)|^{2^k} d\alpha \ll N^{2^k - k + \epsilon}$$

for $\epsilon > 0$.

This motivates to split the interval $[0, 1]$ in two subsets, namely the *major arcs* and *minor arcs*. Although we are working on an interval and not a circle, the terminology comes from the original formulation, which used integration around a circle in the complex plane. For specific examples, one can explicitly define the major and minor arcs and prove with Weyl's inequality and Hua's lemma that the contribution of the minor arcs is small, so that the minor arcs can be regarded as error term. This is often the most difficult step of the Hardy-Littlewood circle method. In most proofs using the method, bounding the order of the major arcs is done in a couple of straightforward steps, while bounding the minor arcs can take pages and requires some high level techniques.

11.2 Comparison to Section 9

When we are summarizing the steps taken in the proofs of Section 9, we would say we translated the counting problem of the inequality

$$|G_k(\mathbf{x}) - \alpha_3 x_3^l| < \delta$$

into bounding an exponential sum

$$f_2(\alpha_3) = \frac{1}{T} \int_{\mathbb{R}} \widehat{\omega_0}\left(\frac{t}{T}\right) F_1(t)\overline{F_2(lt)} e^{-it\log(\alpha_3)} dt$$

as given in (9.7). We divided $f_2(\alpha_3)$ into two parts;

$$f_3(\alpha_3) = \frac{1}{T} \int_{\mathbb{R}} \widehat{\omega_0}\left(\frac{t}{\sqrt{N}}\right) F_1(t)\overline{F_2(lt)} e^{-it\log(\alpha_3)} dt$$

and

$$f_4(\alpha_3) = \frac{1}{T} \int_{\mathbb{R}} \left(\widehat{\omega_0}\left(\frac{t}{T}\right) - \widehat{\omega_0}\left(\frac{t}{\sqrt{N}}\right)\right) F_1(t)\overline{F_2(lt)} e^{-it\log(\alpha_3)} dt.$$

The bound for $f_3(\alpha_3)$ was easily found and is considered as the main term; we could therefore consider this as the *major arcs*. Second, finding a bound for $f_4(\alpha_3)$ was significantly harder. The rest of the proof consisted therefore of finding a bound for this counting function. We concluded that the order of $f_4(\alpha_3)$ was smaller than $f_3(\alpha_3)$, hence $f_4(\alpha_3)$ can be seen as the *minor arcs*.

A natural question arises: if the two methods look so similar, then why did we not use the Hardy-Littlewood circle method? This comes from the fact that the Hardy-Littlewood method uses the counting function $N(f;B)$ which counts rational points *on* an algebraic variety, whereas our problem requires to count rational points *close to* an algebraic variety.

In addition, the function we are looking at in Section 9 is not a polynomial in $Z[x_1, x_2, x_3]$, like f is in $N(f;B)$. The function in inequality (9.1) has irrational coefficients. Therefore, the crucial identity (11.1) for the Hardy-Littlewood circle method is not applicable to finding solutions to the problem in Section 9.

Lastly, the Hardy-Littlewood circle method requires n to be very large in comparison with the degree d of the polynomial. In Section 9, the degree is allowed to be larger than three, resulting in a degree higher than the number of variables. The Hardy-Littlewood circle method is therefore not applicable in the main problem of this thesis.

These three arguments show why we did not apply the Hardy-Littlewood circle method. Instead, the well-thought-out ideas are being used in a different way such that it becomes applicable for functions with irrational coefficients, that are considered close to an algebraic variety.

Bibliography

[1] Birch, B. J. (1962). Forms in many variables. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 265(1321):245–263.

[2] Blomer, V., Bourgain, J., Radziwiłł, M., and Rudnick, Z. (2017). Small gaps in the spectrum of the rectangular billiard. *Annales scientifiques de l'École normale supérieure*, 50(5):1283–1300.

[3] Bourgain, J. (2014). Decoupling, exponential sums and the riemann zeta function. *Journal of the American Mathematical Society*, 30.

[4] Bourgain, J. (2016a). On pair correlation for generic diagonal forms.

[5] Bourgain, J. (2016b). A quantitative oppenheim theorem for generic diagonal quadratic forms. *Israel Journal of Mathematics*, 215(1):503–512.

[6] Browning, T. (2009). *Quantitative Arithmetic of Projective Varieties.* Progress in Mathematics. Birkhäuser Basel.

[7] Brüdern, J. (2013). *Einführung in die analytische Zahlentheorie.* Springer-Lehrbuch. Springer Berlin Heidelberg.

[8] Cohen, H. and Strömberg, F. (2017). *Modular Forms.* Graduate Studies in Mathematics. American Mathematical Society.

[9] Davenport, H. and Browning, T. D. (2005). *Analytic Methods for Diophantine Equations and Diophantine Inequalities.* Cambridge Mathematical Library. Cambridge University Press.

[10] Davenport, H. and Heilbronn, H. (1946). On indefinite quadratic forms in five variables. *Journal of the London Mathematical Society*, 1(3):185–193.

[11] Davenport, H. and Montgomery, H. (1980). *Multiplicative number theory.* Graduate texts in mathematics. Springer-Verlag.

[12] Erdos, P. and Turán, P. (1948). On a problem in the theory of uniform distribution i, ii. *Indag. Math*, 10:370–378.

[13] Eskin, A., Margulis, G., and Mozes, S. (1998). Upper bounds and asymptotics in a quantitative version of the oppenheim conjecture. *Annals of Mathematics*, 147(1):93–141.

[14] Gerstein, L. (2008). *Basic Quadratic Forms.* Graduate studies in mathematics. American Mathematical Society.

[15] Granville, A. and Rudnick, Z. (2007). *Equidistribution in Number Theory, An Introduction.* Nato Science Series II:. Springer Netherlands.

[16] Hardy, G. H. and Littlewood, J. E. (1923). The Approximate Functional Equation in the Theory of the Zeta-Function, with Applications to the Divisor-Problems of Dirichlet and Piltz. *Proceedings of the London Mathematical Society*, s2-21(1):39–74.

[17] Huang, J.-J. (2015). Rational points near planar curves and diophantine approximation. *Advances in Mathematics*, 274:490–515.

[18] Huang, J.-J. (2017). The density of rational points near hypersurfaces. arXiv: 1711.01390.

[19] Iwaniec, H., Kowalski, E., and Society, A. M. (2004). *Analytic Number Theory*. American Mathematical Society Colloquium Publications. American Mathematical Society.

[20] Kronecker, L. (1884). *Näherungsweise ganzzahlige auflösung linearer gleichungen*.

[21] Lang, S. (2013). *Complex Analysis*. Graduate Texts in Mathematics. Springer New York.

[22] Lévy, P. (1939). Sur la division d'un segment par des points choisis au hasard. *CR Acad. Sci. Paris*, 208:147–149.

[23] Lindenstrauss, E. and Margulis, G. (2014). Effective estimates on indefinite ternary forms. *Israel Journal of Mathematics*, 203:445–499.

[24] Margulis, G. A. (1987). Formes quadratiques indéfinies et flots unipotents sur les espaces homogenes. *CR Acad. Sci. Paris Sér. I Math*, 304(10):249–253.

[25] Margulis, G. A. (1989). Discrete subgroups and ergodic theory. In *Number theory, trace formulas and discrete groups*, pages 377–398. Elsevier.

[26] Montgomery, H., Vaughan, R., and Press, C. U. (2007). *Multiplicative Number Theory I: Classical Theory*. Cambridge Studies in Advanced Mathematics. Cambridge University Press.

[27] Nathanson, M. (2013). *Additive Number Theory The Classical Bases*. Graduate Texts in Mathematics. Springer New York.

[28] Oppenheim, A. (1929). The minima of indefinite quaternary quadratic forms. *Proceedings of the National Academy of Sciences of the United States of America*, 15(9):724.

[29] Royden, H. (1988). *Real Analysis*. Mathematics and statistics. Macmillan.

[30] Rudnick, Z. (2017). A metric theory of minimal gaps. *Mathematika*, 64:628–636.

[31] Salberger, P. (2010). Counting rational points on projective varieties. *preprint*.

[32] Sarnak, P. (1996). Values at integers of binary quadratic forms. *Harmonic analysis and number theor*, 21:181–203.

[33] Schindler, D. (2018). Diophantine inequalities for generic ternary diagonal forms. arXiv: 1801.03735.

[34] Serre, J. (2013). *Lectures on the Mordell-Weil Theorem*. Aspects of Mathematics. Vieweg+Teubner Verlag.

[35] Silverman, J. (2009). *The Arithmetic of Elliptic Curves.* Graduate Texts in Mathematics. Springer New York.

[36] Spivak, M. (1965). *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus.* Advanced book program. Avalon Publishing.

[37] Stein, E., Murphy, T., and Press, P. U. (1993). *Harmonic Analysis: Real-variable Methods, Orthogonality, and Oscillatory Integrals.* Monographs in harmonic analysis. Princeton University Press.

[38] Stein, E. and Shakarchi, R. (2010). *Complex Analysis.* Princeton lectures in analysis. Princeton University Press.

[39] Stein, E. and Shakarchi, R. (2011). *Fourier Analysis: An Introduction.* Princeton lectures in analysis. Princeton University Press.

[40] Stephens, L. (2006). *Schaum's Outline of Theory and Problems of Beginning Statistics.* Schaum's outline series. McGraw-Hill.

[41] Tao, T. (2012). *Higher Order Fourier Analysis.* Graduate studies in mathematics. American Mathematical Society.

[42] Titchmarsh, E. C. (1934). On epstein's zeta-function. *Proceedings of the London Mathematical Society*, s2-36(1):485–500.

[43] Titchmarsh, L., Titchmarsh, E., Titchmarsh, E., Heath-Brown, D., and Press, O. U. (1986). *The Theory of the Riemann Zeta-function.* Oxford science publications. Clarendon Press.

[44] Walters, P. (2000). *An Introduction to Ergodic Theory.* Graduate Texts in Mathematics. Springer New York.

[45] Weyl, H. (1916). Über die gleichverteilung von zahlen mod. eins. *Mathematische Annalen*, 77(3):313–352.

[46] Yakubovich, S. (1996). *Index Transforms.* World Scientific.

[47] Zeidler, E. (2016). *Quantum Field Theory I: Basics in Mathematics and Physics: A Bridge between Mathematicians and Physicists.* Springer Berlin Heidelberg.