TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

DISSERTATION

General Gelfond problems for numeration systems

Allgemeine Gelfondsche Probleme in Zahlensystemen

Problèmes de Gelfond généralisés pour les systèmes de numération

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter der Leitung von

Univ.Prof. Michael Drmota Institut für Diskrete Mathematik und Geometrie Technische Universität Wien	und	Pr. Joël Rivat Institut de Mathématiques de Luminy Université de la Méditerranée

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

Dipl.-Ing. Johannes Morgenbesser

Université de la Méditerranée Aix-Marseille II
Faculté des sciences de Luminy
Institut de Mathématiques de Luminy

THÈSE

présentée pour obtenir le grade de

DOCTEUR D'AIX-MARSEILLE UNIVERSITÉ

Spécialité : Mathématiques

par

Johannes MORGENBESSER

sous la direction de

Michael DRMOTA et Joël RIVAT

General Gelfond problems for numeration systems

Problèmes de Gelfond généralisés pour les systèmes de numération

Allgemeine Gelfondsche Probleme in Zahlensystemen

Diese Arbeit wurde im Rahmen eines gemeinsam betreuten Promotionsverfahrens

basierend auf der Vereinbarung

Convention de co-tutelle de thèse

zwischen der
Technischen Universität Wien
und der
Université de la Méditerranée Aix-Marseille II

ausgeführt.

Ce travail est présenté dans le cadre d'une

Convention de co-tutelle de thèse

entre
l'Université de la Méditerranée Aix-Marseille II
et
l'Université Technique de Vienne (Technische Universität Wien).

Danksagung

Mein größter Dank gilt meinen beiden Betreuern Michael Drmota und Joël Rivat. Sie opferten viel Zeit für mich, gaben mir wertvolle Hinweise und Ratschläge, unterstützten mich in meinem Handeln und standen immer hinter mir. Ich konnte durch ihre Hilfe viele neue Einsichten gewinnen und lernte durch sie die Welt der Forschung kennen.

Insbesondere möchte ich mich bei Michael Drmota bedanken, dass er in seinen Vorlesungen mein Interesse an der Zahlentheorie weckte und mir nicht nur das Thema dieser Dissertation schmackhaft machte, sondern mich auch mit den derzeit führenden Forschern in diesem Gebiet zusammenbrachte. Joël Rivat möchte ich für die unzähligen mathematischen und nicht-mathematischen Gespräche danken sowie dafür, dass ich bei meinen Aufenthalten in Marseille auch privat immer willkommen war.

Da ich meine Dissertation im Rahmen einer *Cotutelle de thèse* schrieb, möchte ich mich bei der Technischen Universität Wien und der Université de la Méditerranée sowie bei all jenen bedanken, die dies ermöglichten.

Zum Entstehen dieser Arbeit trugen auch besonders Christian Mauduit und Thomas Stoll bei, welche stets Interesse am Fortschritt meiner Dissertation zeigten. Christian Mauduit möchte ich für die inspirierenden Diskussionen und Anregungen danken sowie Thomas Stoll für das Korrektur Lesen eines Großteils meines Manuskripts. Darüber hinaus konnte ich mich jederzeit mit noch so kleinen (und oftmals trivialen) Fragen an Thomas wenden und bekam immer eine hilfreiche Antwort.

Jean-Marc Deshouillers möchte ich danken, dass er sich bereit erklärt hat, meine Dissertation zu begutachten und Mitglied des Prüfungssenats zu sein. Peter Grabner sei gedankt, dass er ebenfalls akzeptiert hat, meine Dissertation zu begutachten, sowie Rainer Mlitz, dass er den Vorsitz im Prüfungssenat übernimmt.

Weiters bedanke ich mich bei meinen Kollegen Jean-François Bertazzon, Vincent Delecroix, Veronika Kraus, Georg Seitz, Tarek Sellami, Zaid Shawket und Elise Vaslet, dass sie für organisatorische, mathematische und private Fragen immer offen und für knifflige Rätsel immer zu haben waren. Claudia Scheimbauer, Dominik Stürzer und Marie-Claire Fournier schulde ich besonderen Dank für ihre Hinweise zu meinen englischen und französischen Texten. Ich möchte mich auch bei Christoph Haberl, Bruno Martin, Christian Steineder, Wolfgang Steiner und Jörg Thuswaldner für ihr Interesse und ihre Hilfe bedanken.

Großer Dank gilt dem Österreichischen Wissenschaftsfond FWF, der meine Arbeit an der Technischen Universität Wien finanziert hat, sowie dem Institut de Mathématiques de Luminy IML, welches mir immer beste Arbeitsbedingungen in Marseille zur Verfügung gestellt hat.

Schließlich möchte ich mich bei meiner Familie für ihren Rückhalt bedanken sowie bei Cornelia Spreitzer für ihre tägliche Unterstützung aber auch für ihr Verständnis, wenn ich gedankenversunken in einer anderen Welt zu sein schien. Danke!

Johannes Morgenbesser

Remerciements

Je voudrais tout d'abord exprimer mes plus vifs remerciements à mes directeurs de thèse Joël Rivat et Michael Drmota. Ils ont pris beaucoup de temps pour moi, j'ai obtenu des renseignements et conseils précieux et ils ont su me laisser une grande liberté dans ma recherche. Leur soutien constant m'ont permis d'acquérir des connaissances nouvelles et de découvrir le monde de la recherche.

En particulier, je remercie Michael Drmota, qui a éveillé mon intérêt de la théorie des nombres dans les cours de mathématiques et spécialement du sujet de cette thèse. De plus, il m'a présenté aux chercheurs les plus renommés dans cette branche. J'exprime ma profonde gratitude à Joël Rivat pour les conversations innombrables de nature mathématiques et non-mathématiques ainsi que pour l'hospitalité et le fait, que j'ai été toujours le bienvenu chez lui dans mes séjours à Marseille.

J'ai pu faire ce travail dans le cadre d'une cotutelle de thèse. Je tiens à exprimer ma reconnaissance à l'Université de la Méditerranée et à l'Université Technique de Vienne qui m'ont offert cette possibilité ainsi qu'à toutes les personnes qui en ont facilité la mise en œuvre.

Christian Mauduit et Thomas Stoll ont manifesté un intérêt constant pour le développement de ce travail. Merci à Christian Mauduit pour de nombreuses discussions et suggestions qui ont nourri mon inspiration, ainsi qu'à Thomas Stoll pour avoir corrigé une grande part des épreuves. J'ai pu recourir à Thomas pour poser des petites questions (assez souvent triviales) et j'ai toujours obtenu une réponse pertinente.

Je suis honoré que Jean-Marc Deshouillers ait accepté d'être rapporteur pour cette thèse, ainsi que de faire partie du jury. Un grand merci également à Peter Grabner pour avoir accepté d'être rapporteur et à Rainer Mlitz pour avoir accepté d'être président du jury.

Je tiens à saluer mes collègues Jean-François Bertazzon, Vincent Delecroix, Veronika Kraus, Georg Seitz, Tarek Sellami, Zaid Shawket et Elise Vaslet à qui j'ai toujours pu poser des questions administratives, mathématiques et privées. Merci à Claudia Scheimbauer, Dominik Stürzer et Marie-Claire Fournier pour leur renseignements concernant les textes en anglais et en français. Je remercie vivement Christoph Haberl, Bruno Martin, Christian Steineder, Wolfgang Steiner et Jörg Thuswaldner pour leur intérêt et leur aide.

J'adresse un remerciement tout particulier à la fondation autrichienne des sciences (Österreichischen Wissenschaftsfond FWF), qui a financé mon travail à l'Université Technique de Vienne. Une grand merci aussi à l'Institut de Mathématiques de Luminy IML pour avoir mis des bonnes conditions de travail à ma disposition.

Enfin, merci à toute ma famille pour leur soutien et à Cornelia Spreitzer pour son support quotidien mais aussi pour sa compréhension lorsque mes pensées s'évadaient dans un autre monde. Merci!

Johannes Morgenbesser

Contents

Short abstract in German, French, and English

Kurzfassung

Im Jahre 1968 zeigte Gelfond erstmals Gleichverteilungsresultate der Ziffernsumme in allgemeinen arithmetischen Progressionen und stellte weiters drei Fragen. Die erste handelt von der gemeinsamen Verteilung der Ziffernsummen bezüglich verschiedener Basen und wurde 1999 von Kim gelöst. Die zweite und dritte Frage beschäftigen sich mit der Ziffernsumme von Primzahlen und polynomialen Teilfolgen. Mauduit und Rivat lösten vor kurzem diese Probleme für Prim- und Quadratzahlen.

Die vorliegende Dissertation beschäftigt sich mit allgemeinen Gelfondschen Ziffernproblemen und verwandten Fragestellungen. Der erste Teil befasst sich mit der Lösung der Entsprechung des Gelfondschen Problems für Primzahlen in den Gaußschen Zahlen, wobei ein Resultat von Drmota, Rivat und Stoll erweitert wird. Ferner werden spezielle Følner-Folgen (sogenannte κ-$\mathbb{Z}[i]$ Folgen) in den Gaußschen Zahlen eingeführt und Eigenschaften der Ziffernsumme von Quadraten gezeigt. Des Weiteren werden verallgemeinerte Thue-Morse-Folgen in kompakten Gruppen behandelt. Insbesondere wird gezeigt, dass quadratische Teilfolgen solcher Folgen gleichverteilt bezüglich eines Maßes sind, dessen Radon-Nikodym-Dichte mithilfe darstellungstheoretischer Methoden beschrieben werden kann. Als Anwendung werden Häufigkeiten von Buchstaben in Teilfolgen von invertierbaren automatischen Folgen betrachtet. Um das Studium von allgemeinen automatischen Folgen zu ermöglichen, werden matrizenwertige q-multiplikative Funktionen definiert und untersucht. Ferner wird ein Gleichverteilungsresultat für die Ziffernsumme der Folge $(\lfloor n^c \rfloor)_{n\in\mathbb{N}}$ für $c \in \mathbb{R}^+ \setminus \mathbb{N}$ gezeigt, wobei beim Beweis bekannte Exponentialsummenabschätzungen und kürzlich entwickelte Methoden der Fourier-Analyse vereint werden. Schließlich wird eine von Bassily und Kátai ausgearbeitete und von Drmota, Mauduit und Rivat weiterentwickelte Methode verallgemeinert, um lokale Verteilungsresultate auf gewisse Exponentialsummenabschätzungen zurückzuführen. Als Anwendung dieser Methode wird gezeigt, dass die Ziffernsumme von Quadraten in den Gaußschen Zahlen und die Ziffernsumme der Folge $(\lfloor n^c \rfloor)_{n\in\mathbb{N}}$ für $c \in \mathbb{R}^+ \setminus \mathbb{N}$ in den natürlichen Zahlen einem solchen lokalen Verteilungsresultat genügen.

Résumé

En 1968, Gelfond a montré que la fonction somme des chiffres est équirépartie dans les progressions arithmétiques et il a posé trois questions fondamentales. La première concerne la distribution conjointe de la somme des chiffres par rapport à des bases différentes et elle a été résolue par Kim en 1999. La deuxième et la troisième question s'occupent des propriétés de la somme des chiffres pour les nombres premiers et les suites polynomiales. Récemment, Mauduit et Rivat ont réussi à résoudre complètement ces questions dans le cas des suites des nombres premiers et des carrés.

Cette thèse s'occupe des problèmes de Gelfond généralisés et des questions apparentées concernant la somme des chiffres. La première partie de ce travail traite la solution du problème de Gelfond correspondant pour les nombres premiers dans l'anneau des entiers de Gauss où un résultat de Drmota, Rivat et Stoll est amélioré. De plus, on introduit des suites de Følner spéciales (appelées suites de κ-$\mathbb{Z}[i]$) et on montre un résultat de la répartition de la somme des chiffres des carrés. Par ailleurs on traite des suites de Thue-Morse généralisées dans un groupe topologique compact. En particulier, on démontre que les sous-suites des carrés sont équiréparties par rapport à une mesure, dont la dérivée de Radon-Nikodym peut être décrite par la théorie des représentations de groupes. Comme application, on considère des fréquences de lettres dans certaines sous-suites des suites automatiques inversibles. Pour faciliter l'étude des suites automatiques générales, on définit et analyse des fonctions q-multiplicatives généralisées aux matrices complexes. De plus, un résultat de la répartition pour la somme des chiffres de la suite $(\lfloor n^c \rfloor)_{n\in\mathbb{N}}$ avec $c \in \mathbb{R}^+ \setminus \mathbb{N}$ est montré en combinant des estimations des sommes d'exponentielles connues et des méthodes de l'analyse harmonique développées récemment. Enfin, on généralise une méthode développée par Bassily et Kátai et perfectionnée par Drmota, Mauduit et Rivat qui permet d'obtenir des résultats de la répartition locale en étudiant certaines sommes d'exponentielles. Comme application de cette méthode on montre que la somme des chiffres des carrés dans les entiers de Gauss et la somme des chiffres de la suite $(\lfloor n^c \rfloor)_{n\in\mathbb{N}}$ avec $c \in \mathbb{R}^+ \setminus \mathbb{N}$ dans les entiers naturels vérifient un tel résultat local.

Abstract

In 1968, Gelfond proved that the sum-of-digits function is uniformly distributed in arithmetic progressions. This led him to pose three questions: The first asks for the joint distribution of the sum-of-digits function for different bases and was answered completely by Kim in 1999. The second and third question deal with the sum-of-digits function of primes and polynomial subsequences. Mauduit and Rivat recently solved these problems for prime numbers and squares.

This thesis deals with general Gelfond problems and related questions. The first part of this work contains the solution of the corresponding problem of Gelfond for primes in the ring of Gaussian integers. In particular, it includes an extension of a result of Drmota, Rivat, and Stoll. Furthermore, we introduce certain Følner-sequences (so-called κ-$\mathbb{Z}[i]$ sequences) in the Gaussian integers and show distribution results for the sum-of-digits function of squares. Next, we treat generalized Thue-Morse sequences in compact groups. We show that quadratic subsequences of such sequences are uniformly distributed with respect to a measure ν and we determine its Radon-Nikodym derivative with respect to the Haar measure with the help of representation theory. As an application, we consider the frequency of letters in subsequences of invertible automatic sequences. In order to obtain results for subsequences of general automatic sequences, we introduce and analyze matrix-valued q-multiplicative functions. Furthermore, we show that the sum-of-digits function of the sequence $(\lfloor n^c \rfloor)_{n\in\mathbb{N}}$ for $c \in \mathbb{R}^+ \setminus \mathbb{N}$ is uniformly distributed by combining well-known exponential sum estimates and recently developed methods in harmonic analysis. Finally, we generalize a method that was developed by Bassily and Kátai and refined by Drmota, Mauduit, and Rivat in order to trace the study of local distribution results back to certain exponential sum estimates. As an application of this method, we show that the sum-of-digits function of squares in the Gaussian integers and the sum-of-digits function of the sequence $(\lfloor n^c \rfloor)_{n\in\mathbb{N}}$ for $c \in \mathbb{R}^+ \setminus \mathbb{N}$ in the natural numbers obey a local distribution result.

Chapter 1

Introduction in German, French, and English

1.1 Deutsche Einleitung

Die Mathematik ist die Königin
der Wissenschaft, und die Arithmetik ist
die Königin der Mathematik.

Carl Friedrich Gauß (1777-1855)

Die Zahlentheorie, bei Gauß die höhere Arithmetik genannt, ist eine der ältesten Teilgebiete der Mathematik. Euklid hat bereits vor circa 2300 Jahren gezeigt, dass es unendlich viele Primzahlen gibt. Zahlen haben seit jeher auf Mathematikerinnen und Mathematiker eine große Faszination ausgeübt. Seit ein Großteil der modernen Kryptographie auf zahlentheoretischen Überlegungen beruht, hat die Lehre rund um Prim- und Quadratzahlen einen festen Platz in unser aller Leben eingenommen. Vor allem aber die Einfachheit ihrer Fragen und Schönheit ihrer Resultate hat sie zu einem der blühendsten Gebiete der Mathematik gemacht.

Seit Euklids Zeit hat sich die gesamte Mathematik enorm weiterentwickelt. Viele interessante Resultate konnten insbesondere im Bereich der Zahlentheorie gezeigt werden. Die Zahlentheorie ist aber auch jener Bereich, in dem es noch viele offene Fragen gibt. Die Goldbachsche Vermutung (jede gerade Zahl kann als Summe zweier Primzahlen dargestellt werden) und die Frage, ob unendlich viele Primzahlzwillinge existieren (Primzahlenpaare, deren Differenz 2 ergibt), gehören zu den bekanntesten ungelösten Problemen der Mathematik. Nicht unerwähnt soll auch die Riemannsche Vermutung bleiben, die eine zentrale Position in der analytischen Zahlentheorie einnimmt. Ein offenes und mit meinem Dissertationsthema verwandtes, jedoch extrem schwieriges Problem ist die Frage, ob es unendlich viele Primzahlen der Form $2^n - 1$ (sogenannte Mersennesche Primzahlen) oder der Form $2^{2^n} + 1$ (sogenannte Fermatsche Primzahlen) gibt.

Diese beiden Zahlentypen führen mich zu einem anderen Gebiet der Zahlentheorie. Im binären Zahlensystem bestehen Mersennesche Zahlen nur aus Einser (zum Beispiel

$2^3 - 1 = (111)_2$) und Fermatsche Zahlen aus genau einem Einser am Beginn und einem am Ende ($2^{2^2} + 1 = (10001)_2$). Obwohl sich Menschen schon vor tausenden Jahren über die Darstellung von Zahlen Gedanken machten, wurde unser heutzutage übliches Dezimalsystem (inklusive der Ziffer Null) erst im Mittelalter von Indien über den Arabischen Raum nach Europa gebracht. Das im 20. Jahrhundert so wichtig gewordene und bereits zitierte Binärsystem wurde frühestens im 17. Jahrhundert das erste Mal studiert (zur Geschichte der Zahlentheorie siehe [Rib96] und [Knu81, Kapitel 4]). Es gibt wenig bekannte allgemeine Aussagen, die Zahlen und ihre Ziffern in Beziehung setzen. Obwohl solche Probleme teilweise sehr schwierig sind, wurde in den letzten Jahrzehnten einiges auf diesem Sektor geforscht (siehe [MR10, Kapitel 1]). Dazu beigetragen haben dürfte vor allem auch die Tatsache, dass im Zeitalter des Computers Ziffern immens an Bedeutung gewonnen haben.

Von besonderem Stellenwert, sowohl historisch gesehen als auch bezogen auf meine Dissertation, ist die Ziffernsumme einer Zahl n in Basis q (in Zeichen $s_q(n)$). In Basis 2, in der sie der Anzahl der Einser in binärer Schreibweise entspricht, wurde sie indirekt erstmals von Prouhet (1851) studiert. Zu Beginn des 20. Jahrhunderts verwendete Thue die Folge $(s_2(n) \bmod 2)_{n \geqslant 0}$, beginnend mit den Ziffern

$$0\,1\,1\,0\,1\,0\,0\,1\,1\,0\,0\,1\,0\,1\,1\,0\,1\,0\,0\,1\,0\,1\,1\,0\ldots,$$

um ein kombinatorisches Problem zu lösen. Diese Folge wurde von vielen Mathematikern wiederentdeckt (unter anderem von Morse in differentialgeometrischen Überlegungen) und wird heutzutage meist Thue-Morse-Folge genannt (siehe [Mau01]). In der Mitte des 20. Jahrhunderts folgten erste allgemeine Resultate über die Ziffernsumme. Beteiligte Mathematiker waren unter anderem Bellman, Shapiro, Delange, Kátai und Gelfond (zur Geschichte der Ziffernsumme siehe zum Beispiel [AS03]). Letzterer, der Russe Alexander Osipovich Gelfond (Александр Осипович Гельфонд, 1906 – 1968), trug Wesentliches zum Thema der Verteilung der Ziffernsumme bei.

Im Jahre 1968 zeigte er in seiner richtungsweisenden Arbeit *Sur les nombres qui ont des propriétés additives et multiplicatives données* folgendes Resultat:

Satz (Gelfond, [Gel68]). *Es seien q, m und r positive ganze Zahlen mit $q \geqslant 2$ und $(m, q-1) = 1$. Für alle ganze Zahlen ℓ und a gilt*

$$\#\{1 \leqslant n \leqslant N : n \equiv \ell \bmod r,\ s_q(n) \equiv a \bmod m\} = \frac{N}{mr} + O(N^\lambda),$$

wobei $\lambda < 1$ eine positive, nur von q und m abhängige Konstante ist.

Am Ende seiner Abhandlung warf Gelfond drei Fragen auf, welche als Gelfondsche Ziffernsummenprobleme bezeichnet werden. Als erstes vermutete Gelfond, dass für q_1, q_2, m_1, $m_2 \geqslant 2$ mit $(q_1, q_2) = 1$, $(m_1, q_1 - 1) = 1$ und $(m_2, q_2 - 1) = 1$ eine Zahl $\lambda < 1$ existiert, sodass

$$\#\{1 \leqslant n \leqslant N : s_{q_1}(n) \equiv a_1 \bmod m_1 \text{ und } s_{q_2}(n) \equiv a_2 \bmod m_2\} = \frac{N}{m_1 m_2} + O(N^\lambda).$$

Bésineau konnte wenige Jahre später (1972) ein asymptotisches Resultat zeigen, jedoch ohne Angabe einer Fehlertermabschätzung. Fast 30 Jahre später (1999) löste

Kim Gelfonds erstes Problem, wobei ein allgemeingüliger Beweis für q-additive Funktionen angegeben wurde (siehe [Bés72, Kim99]).

Gelfonds zweites Problem beschäftigt sich mit der Anzahl jener Primzahlen, deren Ziffernsumme in einer vorgegebenen Restklasse liegt. Die digitale Beschaffenheit der Ziffernsumme und die multiplikative Struktur von Primzahlen, welche keine offensichtlichen Gemeinsamkeiten haben, machten dieses zu einem besonders interessanten, bis vor kurzem ungelösten Problem. In ihrer Arbeit *Sur un problème de Gelfond: la somme des chiffres des nombres premiers* zeigten Mauduit und Rivat folgendes Resultat:

Satz (Mauduit und Rivat, [MR10]). *Es seien q und m natürliche Zahlen $\geqslant 2$ und $d = (q-1, m)$. Dann existiert eine Konstante $\sigma_{q,m} > 0$, sodass für alle $a \in \mathbb{Z}$*

$$\#\left\{p \leqslant x : p \textit{ prim und } s_q(p) \equiv a \bmod m\right\} = \frac{d}{m}\pi(x; a, d) + O(x^{1-\sigma_{q,m}}),$$

wobei $\pi(x; a, d)$ die Anzahl der Primzahlen $p \leqslant x$ mit $p \equiv a \bmod d$ bezeichnet. Ferner ist die Folge $(\alpha s_q(p))$, p prim, genau dann gleichverteilt modulo 1, wenn $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Die Ziffernsumme polynomialer Folgen steht im Mittelpunkt Gelfonds dritter Frage. Sei P ein Polynom, welches $P(\mathbb{N}) \subseteq \mathbb{N}$ erfüllt. Wie groß ist die Anzahl jener natürlichen Zahlen n, sodass $s_q(P(n)) \equiv a \bmod m$ gilt? Interessanterweise scheint dies Gelfonds schwierigstes Problem zu sein. Ein Grund hierfür ist, dass Folgen, deren n-te Folgenglieder viel größer sind als n, die Arbeit wesentlich erschweren. Im Falle von Quadratzahlen ($P(n) = n^2$) konnten Mauduit und Rivat Gelfonds drittes Problem vollständig lösen.

Satz (Mauduit und Rivat, [MR09]). *Es seien q und m ganze Zahlen $\geqslant 2$. Definiere $d = (q-1, m)$ und $Q(a, d) = \#\left\{0 \leqslant n < d : n^2 \equiv a \bmod d\right\}$. Dann existiert eine Konstante $\sigma_{q,m} > 0$, sodass für alle $a \in \mathbb{Z}$*

$$\#\left\{n \leqslant x : s_q(n^2) \equiv a \bmod m\right\} = \frac{x}{m} Q(a, d) + O\left(x^{1-\sigma_{q,m}}\right).$$

Ferner ist die Folge $(\alpha s_q(n^2))_{n\in\mathbb{N}}$ genau dann gleichverteilt modulo 1, wenn $\alpha \in \mathbb{R}\setminus\mathbb{Q}$.

Meine Arbeit, welche ich im Folgenden kurz vorstelle, beschäftigt sich mit allgemeinen Gelfondschen Ziffernproblemen und verwandten Fragestellungen. Um die Notation in dieser Einleitung so einfach wie möglich zu halten, werde ich einige Sätze nicht in ihrer vollen Allgemeinheit anführen. Exakte Definitionen und ausführliche Resultate befinden sich in den erwähnten Kapiteln. Da in verschiedenen Bereichen der Arbeit die Notation geringfügig variieren kann ($s_q(n)$ bezeichnet zum Beispiel je nach zugrunde liegendem Zahlenbereich die Ziffernsumme in den natürlichen oder den Gaußschen Zahlen), gehe ich zu Beginn jedes Kapitels kurz auf sie ein. Lediglich die natürlichen Zahlen $\mathbb{N}$ definiere ich hiermit inklusive der Zahl 0, da vor allem die Null eine historische (wenn auch erst sehr spät akzeptierte) und für das Studium der Ziffernsumme wichtige Ziffer beschreibt.

In **Kapitel 2** behandle ich die Ziffernsumme von Primzahlen in den Gaußschen Zahlen. Inspiriert durch die Arbeit von Mauduit und Rivat über die Ziffernsumme von Primzahlen in den natürlichen Zahlen bewiesen Drmota, Rivat und Stoll unter bestimmten Annahmen ein ähnliches Resultat in den Gaußschen Zahlen $\mathbb{Z}[i]$ (im Folgenden bezeichnet $s_q(\,\cdot\,)$ die komplexe Ziffernsumme):

Satz (Drmota, Rivat und Stoll, [DRS08]). *Es sei $q = -a \pm i$ eine Primzahl in $\mathbb{Z}[i]$, $a \geqslant 28$ und $b, g \in \mathbb{Z}$, $g \geqslant 2$. Definiere $d = (g, a^2 + 2a + 2)$ und $\delta = (d, 1 \mp i(a + 1))$, wobei die Wahl des Vorzeichens von der Wahl des Vorzeichens von $q = -a \pm i$ abhängt. Dann existiert eine Konstante $\sigma_{q,g} > 0$, sodass*

$$\#\left\{p \in \mathbb{Z}[i] : p \text{ prim}, |p|^2 \leqslant N, s_q(p) \equiv b \bmod g\right\} = \frac{d}{g}\,\pi_i(N; b, d/\delta) + O\left(N^{1-\sigma_{q,g}}\right)$$

mit $\pi_i(N; b, d/\delta) = \#\{|p|^2 \leqslant N : p \text{ prim}, p \equiv b \bmod d/\delta\}$. Ferner ist die Folge $(\alpha s_q(p))$, p prim, genau dann gleichverteilt modulo 1, wenn $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

In meiner Arbeit entledige ich mich den zusätzlichen, etwas unnatürlichen Annahmen von Drmota et al. um Gelfonds zweites Problem in $\mathbb{Z}[i]$ zu lösen. Insbesondere behandle ich alle möglichen Basen, die eine direkte Verallgemeinerung der Ziffernsumme erlauben, und analysiere Primzahlen in Kreissektoren anstelle von Kreisscheiben. In Anbetracht von Heckes Primzahlensatz (Anzahl der Gaußschen Primzahlen in einem Kreissektor) scheinen diese Primzahlbereiche eine natürliche Wahl zu sein. Im Folgenden bezeichnet $\pi_{\gamma_1,\gamma_2}(N; b, d')$ die Anzahl der Gaußschen Primzahlen, welche $|p|^2 \leqslant N$, $\gamma_1 \leqslant \arg(p) < \gamma_2$ und $p \equiv b \bmod d'$ erfüllen.

Satz. *Es sei $q = -a + i$ oder $q = -a - i$, $a \geqslant 1$ eine Gaußsche Zahl und $0 \leqslant \gamma_1 < \gamma_2 \leqslant 2\pi$. Ferner seien b, $g \in \mathbb{Z}$, $g \geqslant 2$, $d = (g, a^2 + 2a + 2)$ und $d' = (g, q - 1)$. Dann existiert eine Konstante $\sigma_{q,g} > 0$, sodass*

$$\begin{aligned}\#\left\{|p|^2 \leqslant N : p \text{ prim}, \gamma_1 \leqslant \arg(p) < \gamma_2, s_q(p) \equiv b \bmod g\right\}& \\ = \frac{d}{g}\pi_{\gamma_1,\gamma_2}(N; b, d') + O(N^{1-\sigma_{q,g}}).&\end{aligned}$$

Bezeichnet $(p_n)_{n\in\mathbb{N}}$ eine Folge aller Gaußschen Primzahlen mit $|p_{n+1}| \geqslant |p_n|$ und $\gamma_1 \leqslant \arg(p_n) < \gamma_2$, dann ist die Folge $(\alpha s_q(p_n))_{n\in\mathbb{N}}$ genau dann gleichverteilt modulo 1, wenn $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Die verwendeten Beweismethoden reichen unter anderem von einer verallgemeinerten Vaughan-Identität in $\mathbb{Z}[i]$ über Approximationseigenschaften spezieller Funktionen bis hin zu einer van der Corput ähnlichen Ungleichung in $\mathbb{Z}[i]$. Ferner benötige ich komplexe Exponentialsummenabschätzungen und L^1- und L^∞-Normabschätzungen der diskreten Fourier-Transformation der komplexen Ziffernsumme.

Die Ziffernsumme von Quadratzahlen in $\mathbb{Z}[i]$ behandle ich in **Kapitel 3**. Das Hauptaugenmerk wird dabei auf die Wahl der Gaußschen Zahlen gelegt, für die ich eine Analyse des Gelfondschen Problems durchführe. Eine Folge $(\mathcal{D}_N)_{N\in\mathbb{N}}$ von Teilmengen der Gaußschen Zahlen heißt κ-$\mathbb{Z}[i]$-Folge, wenn sie die nachstehenden vier Eigenschaften erfüllt:

(i) $\mathcal{D}_N \subseteq \mathcal{D}_{N+1}$,

(ii) $\mathcal{D}_N \subseteq \{z \in \mathbb{Z}[i] : \max(|\Re(z)|, |\Im(z)|) \leqslant \sqrt{N}\}$,

(iii) es existiert eine Konstante $c > 0$, sodass $cN \leqslant \#\mathcal{D}_N$, und

(iv) $\#\{\mathcal{D}_N \triangle (r + \mathcal{D}_N)\} \ll |r| N^{1-\kappa}$ für alle $r \in \mathbb{Z}[i]$.

Diese so definierten Folgen sind spezielle Følner-Folgen. Beispiele für $\kappa = 1/2$ sind Gaußsche Zahlen, die in einem Quadrat mit Seitenlänge $\sqrt{N}$ oder einer Kreisscheibe mit Radius $\sqrt{N}$ liegen (siehe Abbildung 1.1). Interessanterweise kann man zeigen, dass über konvexe Mengen definierte Folgen, welche (i), (ii) und (iii) erfüllen, bereits $1/2$-$\mathbb{Z}[i]$ Folgen sind. Das Hauptresultat dieses Kapitels lautet folgendermaßen, wobei $Q(b,d)$ die Anzahl jener Gaußschen Zahlen z in einem vollständigen Restklassensystem modulo d bezeichnet, deren Quadrate z^2 die Gleichung $z^2 \equiv b \bmod d$ erfüllen:

Satz. *Es sei $q = -a + i$ oder $q = -a - i$, $a \geqslant 1$ eine Gaußsche Zahl, sodass jeder Primteiler von q einen Absolutbetrag $\geqslant \sqrt{689}$ hat. Ferner sei $b, g \in \mathbb{Z}$, $g \geqslant 2$ und $(\mathcal{D}_N)_{N\in\mathbb{N}}$ eine κ-$\mathbb{Z}[i]$-Folge mit $0 < \kappa \leqslant 1/2$. Definiere $d = (g, q-1)$. Dann existiert eine Konstante $\sigma_{q,g,\kappa} > 0$, sodass*

$$\#\{z \in \mathcal{D}_N : s_q(z^2) \equiv b \bmod g\} = \frac{\#\mathcal{D}_N}{g} Q(b,d) + O\left(N^{1-\sigma_{q,g,\kappa}}\right).$$

Bezeichnet $(z_n)_{n\in\mathbb{N}}$ eine geordnete Folge aller Gaußschen Zahlen mit $|z_{n+1}| \geqslant |z_n|$, dann ist die Folge $(\alpha s_q(z_n^2))_{n\in\mathbb{N}}$ genau dann gleichverteilt modulo 1, wenn $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

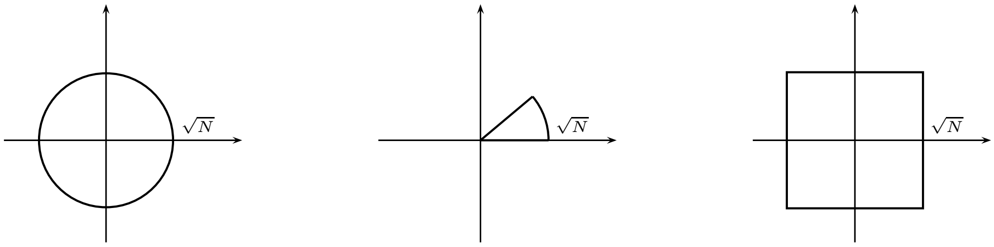


Abbildung 1.1: κ-$\mathbb{Z}[i]$ Folgen mit $\kappa = 1/2$

Ähnlich wie bei den Primzahlen verwende ich wieder eine van der Corput ähnliche Ungleichung. Das besondere an dieser ist, dass sie das Studium der Ziffernsumme in κ-$\mathbb{Z}[i]$-Folgen ermöglicht. Ferner gebe ich eine Formel für spezielle Gauß-Summen in $\mathbb{Z}[i]$ an. Diese sowie die Tatsache, dass die Addition in $\mathbb{Z}[i]$ durch einen Automaten realisiert werden kann, ermöglicht die Verwendung der diskreten Fourier-Transformation der Ziffernsumme. Exponentialsummenabschätzungen, wobei Resultate von Gittenberger und Thuswaldner verbessert werden, sowie die Anwendung der in Kapitel 2 gezeigten Ergebnisse über die Fourier-Transformation schließen den Beweis ab.

In **Kapitel 4** verallgemeinere ich die Ziffernsumme zu einer gruppenwertigen Funktion. Es seien $g_0 = e, \ldots, g_{q-1}$ Elemente einer kompakten Gruppe, wobei $q \geqslant 2$

und e das neutrale Element bezeichnet. Ferner sei G der Abschluss der von den Elementen $g_0, \ldots, g_{q-1}$ erzeugten Untergruppe. Dann heißt die Folge

$$T(n) = g_{\varepsilon_0(n)} g_{\varepsilon_1(n)} \cdots g_{\varepsilon_{\ell-1(n)}}$$

eine verallgemeinerte Thue-Morse-Folge. Hier bezeichnen $\varepsilon_{\ell-1(n)}, \ldots, \varepsilon_{0(n)}$ die Ziffern der Zahl n in Basis q. Sie entspricht einer vollständigen q-multiplikativen G-wertigen Funktion. Wählt man $g_0 = 0$ und $g_1 = 1$ in $\mathbb{Z}/2\mathbb{Z}$ (mit der Addition als Operation), dann entspricht dies genau der klassischen Thue-Morse-Folge.

Satz. *Es sei $(T(n))_{n\in\mathbb{N}}$ eine verallgemeinerte Thue-Morse-Folge. Ferner seien $a \geqslant 1$ und $b \geqslant 0$ ganze Zahlen. Dann existieren absolut stetige Maße ν_1 und ν_2, sodass $T(an+b)_{n\in\mathbb{N}}$ ν_1-gleichverteilt in G sowie $T(n^2)_{n\in\mathbb{N}}$ ν_2-gleichverteilt in G ist.*

Die Darstellungstheorie von kompakten Gruppen ermöglicht es, diese Maße eindeutig zu bestimmen. Ferner können Kriterien angegeben werden, welche entscheiden, ob die betrachteten Folgen gleichverteilt bezüglich des Haarmaßes sind oder nicht. Darstellungen nehmen auch einen bedeutenden Platz in der Beweisführung ein. Es zeigt sich, dass die Methode von Mauduit und Rivat für die Ziffernsumme von Quadraten bestens an irreduzible und unitäre Darstellungen angepasst werden kann.

Automatische Folgen und Häufigkeiten von Buchstaben in Teilfolgen werden in **Kapitel 5** behandelt. Als erstes zeige ich eine Anwendung des Resultats über verallgemeinerte Thue-Morse-Folgen. Wie sich herausstellt, gibt es einen Zusammenhang zwischen endlichen Gruppen und sogenannten invertierbaren q-automatischen Folgen.

Satz. *Es sei $q \geqslant 2$ und $(u_n)_{n\in\mathbb{N}}$ eine invertierbare q-automatische Folge. Dann existiert die Häufigkeit eines jeden Buchstaben in der Teilfolge $(u_{n^2})_{n\in\mathbb{N}}$.*

Leider erlauben verallgemeinerte Thue-Morse-Folgen nicht, universelle automatische Folgen zu betrachten. Folgender Satz lässt sich jedoch mithilfe einer von Mauduit und Rivat entwickelten Methode und der automatischen Struktur der hier betrachteten Folgen zeigen.

Satz. *Es sei $c \in (1, 7/5)$ und $q \geqslant 2$. Ist $(u_n)_{n\in\mathbb{N}}$ eine q-automatische Folge, dann existiert die logarithmische Häufigkeit eines Buchstaben a in $(u_{\lfloor n^c \rfloor})_{n\in\mathbb{N}}$ und sie entspricht der logarithmischen Häufigkeit von a in $(u_n)_{n\in\mathbb{N}}$. Ferner existiert die Häufigkeit von a in $(u_n)_{n\in\mathbb{N}}$ genau dann, wenn sie in $(u_{\lfloor n^c \rfloor})_{n\in\mathbb{N}}$ existiert.*

Um diesen Satz zu beweisen, verallgemeinere ich ein Resultat von Mauduit und Rivat [MR05, Satz 2] und definiere verallgemeinerte q-multiplikative Funktionen im Raum der quadratischen komplexwertigen Matrizen. Einen zentralen Platz im Beweis nimmt ein von Bombieri und Iwaniec entwickeltes doppeltes großes Sieb ein.

In **Kapitel 6** betrachte ich die Ziffernsummenfunktion in den natürlichen Zahlen von $\lfloor n^c \rfloor$, wobei c eine positive reelle Zahl ungleich einer natürlichen Zahl ist. Diese Problemstellung lehnt sich an Gelfonds dritter Frage an, wobei Polynome durch ähnlich stark wachsende Funktionen ersetzt werden. Mauduit und Rivat [MR95, MR05] studierten q-multiplikative Eigenschaften der Folge $(\lfloor n^c \rfloor)_{n\in\mathbb{N}}$ für $c \in (1, 7/5)$ und konnten insbesondere folgendes Resultat zeigen:

Satz (Mauduit und Rivat, [MR05])**.** *Es sei $c \in (1, 7/5)$ und $q \geqslant 2$. Für alle natürlichen Zahlen a und m mit $m \geqslant 1$ gilt*

$$\lim_{x\to\infty} \frac{1}{x} \# \{n \leqslant x : s_q(\lfloor n^c \rfloor) \equiv a \bmod m\} = \frac{1}{m}.$$

Ferner ist die Folge $(\alpha s_q(\lfloor n^c \rfloor))_{n\in\mathbb{N}}$ für $c \in (1, 7/5)$ und $q \geqslant 2$ genau dann gleichverteilt modulo 1, wenn $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Es scheint sehr schwierig zu sein, für größere c ein ähnlich allgemeines Resultat zu erhalten. Unter Verwendung der diskreten Fourier-Transformation kann ich jedoch zeigen, dass für alle positiven reellen c ungleich einer natürlichen Zahl die Folge $s_q(\lfloor n^c \rfloor)$ für genügend große Basen Gleichverteilungseigenschaften besitzt.

Satz. *Es seien $c \in \mathbb{R}^+ \setminus \mathbb{N}$ und $a, m \in \mathbb{N}$ mit $m \geqslant 1$. Dann existiert eine Konstante $q_0(c) \geqslant 2$, sodass für alle $q \geqslant q_0(c)$*

$$\# \{n \leqslant x : s_q(\lfloor n^c \rfloor) \equiv a \bmod m\} = \frac{x}{m} + O\left(x^{1-\sigma_{q,m,c}}\right)$$

mit einer Konstante $\sigma_{q,m,c} > 0$. Ferner ist die Folge $(\alpha s_q(\lfloor n^c \rfloor))_{n\in\mathbb{N}}$ für alle $q \geqslant q_0(c)$ genau dann gleichverteilt modulo 1, wenn $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Mauduit [Mau01] hat die Vermutung aufgestellt, dass folgendes Resultat für fast alle $c > 1$ und für alle $q, m \geqslant 2$ gilt:

$$\lim_{x\to\infty} \frac{1}{x} \# \{n \leqslant x : s_q(\lfloor n^c \rfloor) \equiv a \bmod m\} = \frac{1}{m}. \tag{1.1}$$

Mein Resultat zeigt zwar nicht diese Behauptung, jedoch lässt es vermuten, dass (1.1) genau dann richtig ist, wenn $c \notin \mathbb{N}$ (es ist leicht zu sehen, dass für natürliche Zahlen $c > 1$ Gleichung (1.1) nicht stimmen kann). Um das angeführte Resultat für möglichst viele Basen zu zeigen, verwende ich zwei verschiedene Beweismethoden. Einerseits kann eine von Deshouillers (siehe [Des72] und [Des73a, Des73b]) verwendete Idee benützt werden, welche für alle $c > 1$ funktioniert. Im Bereich zwischen 1 und 19/11 erlaubt andererseits eine Adaptierung der Methode von Mauduit und Rivat [MR05] ein stärkeres Resultat. Neben der erwähnten Fourier-Transformation kommen in beiden Beweismethoden noch klassische Exponentialsummenabschätzungen à la van der Corput und Vinogradov sowie Approximationseigenschaften der Beurling-Selberg-Funktion zum Einsatz.

Kapitel 7 behandelt lokale Verteilungsresultate der natürlichen und komplexen Ziffernsummenfunktion. Drmota, Mauduit und Rivat konnten eine von Bassily und Kátai ausgearbeitete Methode weiterentwickeln und zeigten folgendes Resultat (μ_q und σ_q sind definiert durch $\mu_q = (q-1)/2$ und $\sigma_q^2 = (q^2-1)/12$):

Satz (Drmota, Mauduit und Rivat, [DMR09])**.** *Es sei $q \geqslant 2$. Dann gilt gleichmäßig in $k > 0$ mit $(k, q-1) = 1$, dass $\# \{p \leqslant x : p \text{ prim}, s_q(p) = k\}$ gegeben ist durch*

$$\frac{q-1}{\varphi(q-1)} \frac{\pi(x)}{\sqrt{2\pi\sigma_q^2 \log_q x}} \left(e^{-\frac{(k-\mu_q \log_q x)^2}{2\sigma_q^2 \log_q x}} + O\left((\log x)^{-1/2+\varepsilon} \right) \right),$$

wobei $\varepsilon > 0$ eine beliebige, aber feste positive Zahl ist, $\varphi(\,\cdot\,)$ die Eulersche φ-Funktion bezeichnet und $\pi(x)$ gleich der Anzahl der Primzahlen kleiner gleich x ist.

Ich verallgemeinere diese Arbeit von Drmota, Mauduit und Rivat, und zeige, dass man unter gewissen Bedingungen eine asymptotische Entwicklung der Zahlen $\#\{n \leqslant x : s_q(g(n)) = k\}$ erhält, wobei $g(n)$ eine Funktion von $\mathbb{N}$ nach $\mathbb{N}$ ist. Weiters behandle ich eine analoge Fragestellung in den Gaußschen Zahlen, wobei ich Ideen von Gittenberger und Thuswaldner weiterentwickle. Schließlich gebe ich einige Beispiele an, welche die Nützlichkeit dieser Überlegungen unterstreichen. Insbesondere erlauben mir die Ergebnisse aus Kapitel 6, folgenden Satz zu zeigen:

Satz. *Es sei $c \in \mathbb{R}^+ \setminus \mathbb{N}$. Dann existiert eine Konstante $q_0(c) \geqslant 2$, sodass für alle $q \geqslant q_0(c)$ und gleichmäßig in $k \in \mathbb{N}$*

$$\begin{aligned}\frac{1}{x}\#\left\{n \leqslant x : s_q(\lfloor n^c \rfloor) = k\right\}\\ = \frac{x}{\sqrt{2\pi\sigma_q^2 c \log_q x}}\left(e^{-\frac{(k-\mu_q c \log_q x)^2}{2\sigma_q^2 c \log_q x}} + O\left(\frac{(\log\log x)^7}{(\log x)^{1/2}}\right)\right).\end{aligned}$$

Unter Verwendung der Resultate aus Kapitel 3 und unter neuerlichem Einsatz einer van der Corput ähnlichen Ungleichung in $\mathbb{Z}[i]$ erhalte ich ebenfalls ein Resultat für die asymptotische Entwicklung der Zahlen $\#\{z \in \mathcal{D}_N : s_q(z^2) = k\}$, wobei $(\mathcal{D}_N)_{n\in\mathbb{N}}$ beliebige $1/2$-$\mathbb{Z}[i]$ Folgen sein können.

Abschließend möchte ich noch einmal die wichtigsten Ergebnisse meiner Dissertation kurz zusammenfassen und Referenzen zu von mir in wissenschaftlichen Journalen eingereichten oder bereits veröffentlichen Arbeiten geben. Um das Gelfondsche Problem für Primzahlen in $\mathbb{Z}[i]$ vollständig zu lösen, behandle ich die diskrete Fourier-Transformation der Ziffernsumme im Detail und verbessere einige bekannte Exponentialsummenabschätzungen in $\mathbb{Z}[i]$ (**Kapitel 2**). In den Gaußschen Zahlen führe ich spezielle Følner-Folgen ein (sogenannte κ-$\mathbb{Z}[i]$ Folgen) und entwickle eine van der Corput ähnliche Ungleichung, um die Ziffernsumme von Quadratzahlen behandeln zu können (**Kapitel 3 und 7**). Der Inhalt dieser Kapitel beruht auf meinem im *Journal of Number Theory* veröffentlichten Artikel [Mor10a] und auf meiner eingereichten Arbeit [Mor10b]. Ferner studiere ich verallgemeinerte Thue-Morse-Folgen in kompakten Gruppen, indem ich eine zur Arbeit von Mauduit und Rivat analoge Methode für Darstellungen entwickle (**Kapitel 4**). Diese Überlegungen beruhen auf der mit Michael Drmota gemeinsam eingereichten und im *Israel Journal of Mathematics* zur Publikation angenommenen Arbeit [DM10]. Als Anwendung berechne ich Häufigkeiten von Buchstaben in quadratischen Teilfolgen von invertierbaren automatischen Folgen. Um das Studium von gewissen Teilfolgen von allgemeinen automatischen Folgen zu ermöglichen, definiere und untersuche ich verallgemeinerte matrizenwertige q-multiplikative Funktionen (**Kapitel 5**). Teile dieser Resultate befinden sich ebenfalls in [DM10]. Schließlich behandle ich noch die Ziffernsumme der Folge $(\lfloor n^c \rfloor)_{n\in\mathbb{N}}$ für $c \in \mathbb{R}^+ \setminus \mathbb{N}$, wobei ich altbekannte Exponentialsummenabschätzungen und kürzlich entwickelte Methoden der Fourier-Analyse vereine (**Kapitel 6 und 7**). Diese Resultate beruhen auf der von mir eingereichten Arbeit [Mor10c].

1.2 Introduction en français

Les mathématiques sont la reine des sciences, et l'arithmétique est la reine des mathématiques.

Carl Friedrich Gauss (1777-1855)

La théorie des nombres (l'arithmétique) est l'une des branches mathématiques les plus anciennes. Euclide a déjà montré il y a plus de 2300 ans qu'il existe une infinité de nombres premiers. De tout temps les mathématiciennes et les mathématiciens ont été fascinés par les nombres entiers. Puisqu'une grande partie de la cryptographie est basée sur des principes de la théorie des nombres, l'étude autour des nombres premiers et des carrés a pris une place très importante dans notre quotidien. La simplicité de ses questions et la beauté de ses résultats ont fait de la théorie des nombres l'un des domaines mathématiques les plus florissants.

Depuis le temps d'Euclide les mathématiques ont énormément évolué. Beaucoup de résultats intéressants ont été montrés notamment dans le domaine de la théorie des nombres. De plus, la théorie des nombres est un domaine où il y a beaucoup de questions ouvertes. La conjecture de Goldbach (chaque nombre pair est la somme de deux nombres premiers) et la question de savoir s'il y a une infinité de nombres premiers jumeaux (des couples de nombres premiers qui ne diffèrent que de 2) font partie des problèmes non résolus et les plus connus. Je voudrais aussi mentionner l'hypothèse de Riemann qui joue un rôle capital dans la théorie analytique des nombres. Un problème ouvert très proche du sujet de ma thèse mais hors d'atteinte est de déterminer s'il y a une infinité de nombres premiers de la forme $2^n - 1$ (nombres premiers de Mersenne) ou de la forme $2^{2^n} + 1$ (nombres premiers de Fermat).

Ce deux types de nombres me conduisent à une branche de la théorie des nombres qui est, dans un certain sens, moins vieille que l'étude des nombres premiers. Bien que les gens réfléchissent déjà depuis plus de mille ans sur le système de numération, notre système décimal commun (avec le chiffre zéro) n'a été apporté qu'au Moyen Âge de l'Inde à l'Europe via le monde arabe. Le système binaire qui est devenu très important dans le XXième siècle, n'a été développé qu'au XVIIième siècle (au sujet de l'histoire de la théorie des nombres et des chiffres voir par exemple [Rib96] et [Knu81, Chapitre 4]).

Ce système me ramène aux nombres de Mersenne et de Fermat. En base 2, l'écriture des nombres de Mersenne ne comporte que des chiffres 1 (par exemple $2^3 - 1 = (111)_2$) et les nombres de Fermat ont exactement un 1 au début et un 1 à la fin de la représentation binaire (par exemple $2^{2^2} + 1 = (10001)_2$). Il y a très peu de résultats généraux pour les nombres qui ont des propriétés spéciales données par leurs chiffres. Bien que les problèmes des chiffres soient en général très difficiles, beaucoup de personnes ont entrepris des recherches durant les dernières décennies dans cette branche (voir par exemple [MR10, Section 1]). C'est peut-être parce que les chiffres ont acquis une importance immense avec le développement des ordinateurs.

Dans ce travail je vais aborder des questions consacrées aux nombres et aux propriétés de leurs chiffres. La somme des chiffres de n en base q (notée $s_q(n)$) joue un rôle très important, non seulement historique, mais aussi pour le contenu de ma thèse. Par exemple, quand on écrit un entier n en base 2, on a besoin de $s_2(n)$ fois le chiffre 1. Dans cette forme, la somme des chiffres est étudiée indirectement par Prouhet en 1851. Au début du XXième siècle Thue a utilisé la suite $(s_2(n) \bmod 2)_{n\geqslant 0}$ qui commence par

$$0\,1\,1\,0\,1\,0\,0\,1\,1\,0\,0\,1\,0\,1\,1\,0\,1\,0\,0\,1\,0\,1\,1\,0\ldots$$

pour résoudre un problème de combinatoire. Cette suite a été redécouverte plusieurs fois par des mathématiciens différents (entre autres Morse dans le domaine de la géométrie différentielle) et aujourd'hui la plupart du temps elle s'appelle la suite de Thue-Morse, voir [Mau01]. Au milieu du XXième siècle les premiers résultats plus généraux ont été obtenus pour la fonction « la somme des chiffres ». Je voudrais mentionner en particulier Bellman, Shapiro, Delange, Kátai et Gelfond (au sujet de l'histoire de la somme des chiffres voir par exemple [AS03]). Le dernier de cette liste, le mathématicien russe Aleksandr Osipovich Gelfond (Александр Осипович Гельфонд, 1906 – 1968) a contribué beaucoup à l'étude de la répartition de la somme des chiffres. En 1968 il a montré dans son travail novateur *« Sur les nombres qui ont des propriétés additives et multiplicatives données »* le résultat suivant :

Théorème (Gelfond, [Gel68]). *Soient q, m et r des entiers positifs tels que $q \geqslant 2$ et $(m, q-1) = 1$. Alors pour tous entiers ℓ et a on a*

$$\operatorname{card}\{1 \leqslant n \leqslant N : n \equiv \ell \bmod r,\ s_q(n) \equiv a \bmod m\} = \frac{N}{mr} + O(N^\lambda)$$

avec une constante $\lambda < 1$ qui ne dépend que de q et m.

À la fin de son article, Gelfond a posé trois problèmes qui sont liés à son théorème. Ensuite, j'aborderai ceux qui sont très centraux dans ma thèse et je citerai quelques résultats connus. Tout d'abord il a conjecturé que pour tous les nombres entiers q_1, q_2, m_1 et m_2 avec $(q_1, q_2) = 1$, $(m_1, q_1 - 1) = 1$ et $(m_2, q_2 - 1) = 1$ il existe une constante $\lambda < 1$ telle que

$$\operatorname{card}\{1 \leqslant n \leqslant N : s_{q_1}(n) \equiv a_1 \bmod m_1 \text{ et } s_{q_2}(n) \equiv a_2 \bmod m_2\} = \frac{N}{m_1 m_2} + O(N^\lambda).$$

Bésineau a montré un peu plus tard (1972) un résultat asymptotique, mais sans estimation du terme d'erreur. Presque 30 ans plus tard (1999) Kim a résolu le premier problème de Gelfond plus généralement pour des fonctions q-additives (voir [Bés72, Kim99]).

Le deuxième problème de Gelfond se consacre au nombre des nombres premiers dont la somme des chiffres se trouve dans une certaine classe modulo m (où m est un entier positif). La propriété digitale de la somme des chiffres et la structure multiplicative des nombres premiers n'ont a priori pas de points communs. Ce problème est d'une difficulté extraordinaire et il n'a pu être résolu que récemment. Dans leur travail *« Sur un problème de Gelfond : la somme des chiffres des nombres premiers »* Mauduit et Rivat ont donné une réponse complète à la question de Gelfond et ils ont montré le résultat suivant :

Théorème (Mauduit et Rivat, [MR10]). *Soient q et m des entiers $\geqslant 2$ et posons $d = (q-1, m)$. Alors il existe une constante $\sigma_{q,m} > 0$ telle que pour tout $a \in \mathbb{Z}$,*

$$\operatorname{card}\{p \leqslant x : p \text{ premier, } s_q(p) \equiv a \bmod m\} = \frac{d}{m}\pi(x; a, d) + O(x^{1-\sigma_{q,m}})$$

où on désigne par $\pi(x; a, d)$ le nombre des nombres premiers $p \leqslant x$ tels que $p \equiv a \bmod d$. De plus, la suite $(\alpha s_q(p))$, p premier, est équirépartie modulo 1 si et seulement si $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Le troisième problème de Gelfond s'occupe des propriétés de la somme des chiffres pour les sous-suites polynomiales. Soit P un polynôme qui vérifie $P(\mathbb{N}) \subseteq \mathbb{N}$. Combien existe-t-il d'entiers n tels que $s_q(P(n)) \equiv a \bmod m$? Curieusement, cette question semble le problème de Gelfond le plus difficile. Une raison pour cette circonstance est le fait que l'étude des propriétés des suites éparses (le n-ième terme de la suite est beaucoup plus grand que n) est très complexe en général. Dans le cas de la suite des carrés ($P(n) = n^2$), Mauduit et Rivat ont réussi à résoudre complètement le troisième problème de Gelfond :

Théorème (Mauduit et Rivat, [MR09]). *Soient q et m des entiers avec $q \geqslant 2$ et posons $d = (q-1, m)$ et $Q(a, d) = \operatorname{card}\{0 \leqslant n < d : n^2 \equiv a \bmod d\}$. Alors il existe une constante $\sigma_{q,m} > 0$ telle que pour tout $a \in \mathbb{Z}$,*

$$\operatorname{card}\{n \leqslant x : s_q(n^2) \equiv a \bmod m\} = \frac{x}{m} Q(a, d) + O\left(x^{1-\sigma_{q,m}}\right).$$

De plus, la suite $(\alpha s_q(n^2))_{n \in \mathbb{N}}$ est équirépartie modulo 1 si et seulement si $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Cette thèse s'occupe des problèmes de Gelfond généralisés. Afin de minimiser les notations dans cette introduction qui a pour but de donner une idée d'ensemble de mon travail, je ne citerai pas certains théorèmes en pleine généralité. Des définitions exactes et des résultats détaillés se trouvent dans les chapitres mentionnés. Puisque la notation peut varier un peu dans des domaines différents ($s_q(n)$ dénote par exemple à la fois la somme des chiffres dans les entiers naturels et celle des entiers de Gauss), je vais introduire en détail la notation appropriée dans chaque chapitre. Seulement je définis ici l'ensemble des entiers naturels $\mathbb{N}$ en incluant le nombre 0, parce que le zéro est un chiffre historique (même s'il n'est accepté que très tard) et est très important pour l'étude de la somme des chiffres.

Dans le **Chapitre 2** je traite la somme des chiffres dans l'anneau des entiers de Gauss. Inspiré par le travail de Mauduit et Rivat concernant la somme des chiffres commune, Drmota, Rivat et Stoll ont montré sous certaines hypothèses un résultat similaire dans l'anneau des entiers de Gauss $\mathbb{Z}[i]$ (dans le théorème suivant, $s_q(\cdot)$ dénote la somme des chiffres pour les entiers de Gauss) :

Théorème (Drmota, Rivat et Stoll, [DRS08]). *Soit $-a+i$ un nombre premier dans $\mathbb{Z}[i]$ tel que $a \geqslant 28$ est un entier et soient $b, g \in \mathbb{Z}$, $g \geqslant 2$. Supposons que $q = -a+i$*

ou $q = -a - i$. Si $d = (g, a^2 + 2a + 2)$ et $\delta = (d, 1 \mp i(a+1))$ (où le choix du signe dépend du signe de $q = -a \pm i$), il existe une constante $\sigma_{q,g} > 0$ telle que

$$\begin{aligned}\operatorname{card}\{p \in \mathbb{Z}[i] : p \textit{ premier, } |p|^2 \leqslant N, s_q(p) \equiv b \bmod g\}& \\ = \frac{d}{g}\, \pi_i(N; b, d/\delta) + O\left(N^{1-\sigma_{q,g}}\right)&\end{aligned}$$

où on désigne par $\pi_i(N; b, d/\delta)$ le nombre des nombres premiers de Gauss $|p|^2 \leqslant N$ tels que $p \equiv b \bmod d/\delta$. De plus, la suite $(\alpha s_q(p))$, p premier, est équirépartie modulo 1 si et seulement si $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Drmota, Rivat et Stoll ont montré ce résultat seulement pour certaines bases très spéciales et pour certains domaines des nombres premiers. Dans ce travail je peux me débarrasser de ces hypothèses et réussis à résoudre complètement le deuxième problème de Gelfond dans $\mathbb{Z}[i]$. En particulier, je traite toutes les bases qui permettent une généralisation directe de la somme des chiffres dans $\mathbb{Z}[i]$ et j'analyse les nombres premiers dans les secteurs circulaires au lieu des disques. Le théorème des nombres premiers de Hecke (qui s'occupe du nombre des nombres premiers de Gauss dans un secteur circulaire) montre que ces domaines sont canoniques. Dans ce qui suit, $\pi_{\gamma_1,\gamma_2}(N; b, d')$ désigne le nombre des nombres premiers de Gauss qui vérifient $|p|^2 \leqslant N$, $\gamma_1 \leqslant \arg(p) < \gamma_2$ et $p \equiv b \bmod d'$.

Théorème. *Soient $q = -a + i$ ou $q = -a - i$, $a \geqslant 1$ et $0 \leqslant \gamma_1 < \gamma_2 \leqslant 2\pi$. Supposons que b, $g \in \mathbb{Z}$, $g \geqslant 2$ et posons $d = (g, a^2 + 2a + 2)$ et $d' = (g, q - 1)$. Alors, il existe une constante $\sigma_{q,g} > 0$ telle que*

$$\begin{aligned}\operatorname{card}\{|p|^2 \leqslant N : p \textit{ premier, } \gamma_1 \leqslant \arg(p) < \gamma_2,\ s_q(p) \equiv b \bmod g\}& \\ = \frac{d}{g} \pi_{\gamma_1,\gamma_2}(N; b, d') + O(N^{1-\sigma_{q,g}}).&\end{aligned}$$

Soit $(p_n)_{n\in\mathbb{N}}$ une suite des nombres premiers de Gauss avec $\gamma_1 \leqslant \arg(p_n) < \gamma_2$ et $|p_{n+1}| \geqslant |p_n|$. Alors la suite $(\alpha s_q(p_n))_{n\in\mathbb{N}}$ est équirépartie modulo 1 si et seulement si $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Pour montrer ce résultat j'utilise et extrapole plusieurs méthodes différentes : en particulier, j'applique une variante généralisée dans $\mathbb{Z}[i]$ de l'identité de Vaughan, des approximations d'une fonction spéciale et une inégalité de type van der Corput dans $\mathbb{Z}[i]$. De plus, je traite des estimations des sommes d'exponentielles complexes et des estimations des normes L^1 et L^∞ de la transformée de Fourier discrète de la somme des chiffres.

Dans le chapitre suivant (**Chapitre 3**) je m'occupe de la somme des chiffres des carrés dans $\mathbb{Z}[i]$. Je précise le choix des entiers de Gauss que je vais analyser. Une suite $(\mathcal{D}_N)_{N\in\mathbb{N}}$ de sous-ensembles de $\mathbb{Z}[i]$ est appelée suite de κ-$\mathbb{Z}[i]$, si elle vérifie les quatre propriétés suivantes :

(i) $\mathcal{D}_N \subseteq \mathcal{D}_{N+1}$,
(ii) $\mathcal{D}_N \subseteq \{z \in \mathbb{Z}[i] : \max\left(|\Re(z)|, |\Im(z)|\right) \leqslant \sqrt{N}\,\}$,

(iii) il existe une constante $c > 0$ telle que $cN \leqslant \operatorname{card} \mathcal{D}_N$, et

(iv) $\operatorname{card}\{\mathcal{D}_N \triangle (r + \mathcal{D}_N)\} \ll |r| N^{1-\kappa}$ pour tout $r \in \mathbb{Z}[i]$.

Notons que ces suites sont des suites spéciales de Følner. Les exemples les plus faciles (pour $\kappa = 1/2$) sont les entiers de Gauss qui sont situés dans un carré de côté $\sqrt{N}$ ou un disque de rayon $\sqrt{N}$, voir Figure 1.2. Curieusement, si une suite de κ-$\mathbb{Z}[i]$ est définie par des ensembles convexes (qui vérifient (i), (ii) et (iii)), on peut montrer facilement qu'elle est automatiquement une suite de $1/2$-$\mathbb{Z}[i]$. Le résultat principal dans ce chapitre est le suivant ($Q(b, d)$ dénote le nombre des entiers de Gauss z dans une classe d'équivalence modulo d avec $z^2 \equiv b \bmod d$) :

Théorème. *Soit $q = -a + i$ ou $q = -a - i$, $a \geqslant 1$ un entier de Gauss tel que chaque facteur premier de q possède une valeur absolue $\geqslant \sqrt{689}$. Supposons que $b, g \in \mathbb{Z}$, $g \geqslant 2$ et que $(\mathcal{D}_N)_{N \in \mathbb{N}}$ soit une suite de κ-$\mathbb{Z}[i]$ avec $0 < \kappa \leqslant 1/2$. Posons $d = (g, q-1)$. Alors, il existe une constante $\sigma_{q,g,\kappa} > 0$ telle que*

$$\operatorname{card}\{z \in \mathcal{D}_N : s_q(z^2) \equiv b \bmod g\} = \frac{\operatorname{card} \mathcal{D}_N}{g} Q(b, d) + O\left(N^{1-\sigma_{q,g,\kappa}}\right).$$

Soit $(z_n)_{n \in \mathbb{N}}$ une suite des entiers de Gauss avec $|z_{n+1}| \geqslant |z_n|$. Alors la suite $(\alpha s_q(z_n^2))_{n \in \mathbb{N}}$ est équirépartie modulo 1 si et seulement si $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

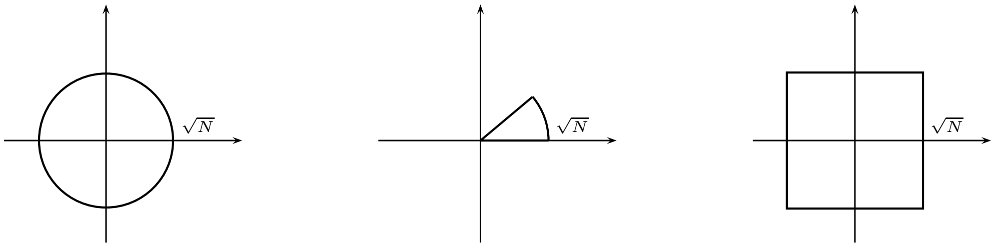


FIG. 1.2 – Suites de κ-$\mathbb{Z}[i]$ avec $\kappa = 1/2$

Par analogie avec les nombres premiers j'utilise une inégalité de type van der Corput généralisée. Mais cette fois, elle me permet de ramener l'étude des suites de κ-$\mathbb{Z}[i]$ à l'étude d'une suite plus simple à utiliser. De plus, je traite les sommes de Gauss dans $\mathbb{Z}[i]$. Ces sommes et le fait que l'addition dans l'anneau des entiers de Gauss peut être réalisée par un automate fini, me permettent d'appliquer la transformée de Fourier. Des estimations précises des sommes d'exponentielles (des résultats de Gittenberger et Thuswaldner sont améliorés) et des résultats au sujet de la transformée de Fourier montrés dans le Chapitre 2 terminent la démonstration du théorème cité.

Ensuite, je généralise dans le **Chapitre 4** la suite de Thue-Morse et je définis une suite dans un groupe topologique compact H. Soit $q \geqslant 2$ et soient $g_0, \ldots, g_{q-1}$ des éléments appartenant à H. Supposons que g_0 est l'élément neutre et notons par G l'adhérence du groupe engendré par les éléments $g_0, \ldots, g_{q-1}$. On appelle

$$T(n) = g_{\varepsilon_0(n)} g_{\varepsilon_1(n)} \cdots g_{\varepsilon_{\ell-1(n)}}$$

$n \geqslant 0$, la suite de Thue-Morse généralisée (où $\varepsilon_{\ell-1(n)}, \ldots, \varepsilon_{0(n)}$ sont les chiffres de n en base q). Elle est une suite complètement q-multiplicative et si on choisit $g_0 = 0$ et $g_1 = 1$ dans $\mathbb{Z}/2\mathbb{Z}$ (où l'addition est l'opération du groupe), on obtient la suite de Thue-Morse classique. Par ailleurs, on peut réaliser la suite $(\alpha s_q(n))_{n\in\mathbb{N}}$, α réel, sur le tore.

Théorème. *Soit $(T(n))_{n\in\mathbb{N}}$ une suite de Thue-Morse généralisée. Supposons que $a \geqslant 1$ et $b \geqslant 0$ sont des entiers. Alors il existe des mesures positives ν_1 et ν_2 qui sont absolument continues par rapport à la mesure de Haar telles que $T(an+b)_{n\in\mathbb{N}}$ est équirépartie par rapport à ν_1 et $T(n^2)_{n\in\mathbb{N}}$ est équirépartie par rapport à ν_2.*

La théorie des représentations des groupes topologiques permet de déterminer complètement ces mesures. De plus, il existe un critère qui permet de décider si la suite considérée est équirépartie par rapport à la mesure de Haar. La théorie des représentations prend une grande place dans toute la démonstration. On peut constater que la méthode de Mauduit et Rivat pour la somme des chiffres des carrés s'adapte très bien à l'étude des représentations irréductibles et unitaires.

Dans le **Chapitre 5** je traite des suites automatiques. En particulier, je considère des fréquences logarithmiques et naturelles des lettres d'une suite q-automatique sur un alphabet Δ. Dans un premier temps, je montre une application des résultats de la suite de Thue-Morse généralisée. Il y a un lien entre les groupes finis et les suites q-automatiques spéciales (suites q-automatiques inversibles, voir Chapitre 5). L'exemple le plus simple de telles suites est la suite de Thue-Morse classique.

Théorème. *Soient $q \geqslant 2$ et $(u_n)_{n\in\mathbb{N}}$ une suite q-automatique inversible sur l'alphabet Δ. Si $a \in \Delta$, alors la lettre a admet une fréquence dans la sous-suite $(u_{n^2})_{n\in\mathbb{N}}$.*

Malheureusement, la suite de Thue-Morse généralisée ne permet pas un traitement des suites automatiques sans restrictions. Mais je peux montrer le théorème suivant avec une méthode développée par Mauduit et Rivat.

Théorème. *Soient $c \in (1, 7/5)$, $q \geqslant 2$ et $(u_n)_{n\in\mathbb{N}}$ une suite q-automatique sur l'alphabet Δ. Supposons que $a \in \Delta$. Alors il existe la fréquence logarithmique de a dans la sous-suite $(u_{\lfloor n^c \rfloor})_{n\in\mathbb{N}}$ et elle est donnée par la fréquence logarithmique de a dans $(u_n)_{n\in\mathbb{N}}$. De plus, il existe la fréquence de a dans $(u_n)_{n\in\mathbb{N}}$ si et seulement s'il existe la fréquence de a dans $(u_{\lfloor n^c \rfloor})_{n\in\mathbb{N}}$.*

En particulier, je généralise un résultat de Mauduit et Rivat (voir [MR05, Théorème 2]) concernant les fonctions q-multiplicatives que j'obtiens dans l'anneau des matrices carrées. Dans le cadre de cette généralisation je définis des fonctions q-multiplicatives dans $\mathbb{C}^{d\times d}$. Le double grand crible de Bombieri et Iwaniec joue un rôle très important pour la démonstration.

Le **Chapitre 6** s'occupe de la somme des chiffres de la suite $(\lfloor n^c \rfloor)_{n\in\mathbb{N}}$, où c est un réel positif et non entier. Ce problème est inspiré par la troisième question de Gelfond (il s'agit de remplacer les polynômes par des fonctions de croissance comparable). Mauduit et Rivat [MR95, MR05] ont étudié des propriétés q-multiplicatives de la suite $(\lfloor n^c \rfloor)_{n\in\mathbb{N}}$ pour $c \in (1, 7/5)$. En particulier, ils ont démontré le résultat suivant :

Théorème (Mauduit et Rivat, [MR05]). *Soient $c \in (1, 7/5)$ et $q \geqslant 2$. Pour tous les entiers a et m avec $m \geqslant 1$ on a*

$$\lim_{x\to\infty} \frac{1}{x} \operatorname{card}\{n \leqslant x : s_q\left(\lfloor n^c \rfloor\right) \equiv a \bmod m\} = \frac{1}{m}.$$

De plus, la suite $(\alpha s_q(\lfloor n^c \rfloor))_{n\in\mathbb{N}}$ est équirépartie modulo 1 si et seulement si $\alpha \in \mathbb{R}\backslash\mathbb{Q}$.

Il semble très difficile de montrer un résultat similaire pour un nombre c plus grand. Dans la suite je montre cependant avec l'utilisation de la transformée de Fourier discrète de la somme des chiffres que pour q assez grand et pour tout réel positif non entier c la suite $s_q(\lfloor n^c \rfloor)$ est bien répartie dans les progressions arithmétiques.

Théorème. *Soient $c \in \mathbb{R}^+ \setminus \mathbb{N}$ et $a, m \in \mathbb{N}$ avec $m \geqslant 1$. Alors il existe une constante $q_0(c) \geqslant 2$ telle que pour tout $q \geqslant q_0(c)$ on a*

$$\operatorname{card}\{n \leqslant x : s_q\left(\lfloor n^c \rfloor\right) \equiv a \bmod m\} = \frac{x}{m} + O\left(x^{1-\sigma_{q,m,c}}\right),$$

avec une constante $\sigma_{q,m,c} > 0$. De plus, pour $q \geqslant q_0(c)$ la suite $(\alpha s_q(\lfloor n^c \rfloor))_{n\in\mathbb{N}}$ est équirépartie modulo 1 si et seulement si $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Mauduit a conjecturé dans [Mau01], que le résultat suivant est valide pour presque tous les réels $c > 1$ et pour tous $q, m \geqslant 2$:

$$\lim_{x\to\infty} \frac{1}{x} \operatorname{card}\{n \leqslant x : s_q(\lfloor n^c \rfloor) \equiv a \bmod m\} = \frac{1}{m}. \tag{1.2}$$

Mon résultat ne montre pas cette assertion, mais je conjecture que (1.2) est vrai si est seulement si c est non entier (on vérifie facilement que (1.2) n'est pas vrai pour un entier $c > 1$). Pour montrer le théorème cité pour la plupart des bases, j'emploie deux méthodes différentes. Deshouillers (voir [Des72] et [Des73a, Des73b]) utilise une idée qui fonctionne pour tout $c > 1$, mais entre 1 et 19/11 une adaptation de la méthode de Mauduit et Rivat permet un meilleur résultat. À part la transformée de Fourier j'utilise dans les deux méthodes des estimations des sommes d'exponentielles classiques (établies par van der Corput et Vinogradov) ainsi que des approximations d'une fonction à variation bornée utilisant la fonction de Beurling-Selberg.

Enfin, dans le **Chapitre 7** je traite la répartition locale de la somme des chiffres des entiers. Drmota, Mauduit et Rivat ont perfectionné une méthode développée par Bassily et Kátai et ils ont montré le théorème suivant (μ_q et σ_q sont définis par $\mu_q = (q-1)/2$ et $\sigma_q^2 = (q^2-1)/12$) :

Théorème (Drmota, Mauduit et Rivat, [DMR09]). *Soit $q \geqslant 2$. On a uniformément pour $k > 0$ avec $(k, q-1) = 1$, que $\operatorname{card}\{p \leqslant x : p$ premier, $s_q(p) = k\}$ est donné par*

$$\frac{q-1}{\varphi(q-1)} \frac{\pi(x)}{\sqrt{2\pi\sigma_q^2 \log_q x}} \left(e^{-\frac{(k-\mu_q \log_q x)^2}{2\sigma_q^2 \log_q x}} + O\left((\log x)^{-1/2+\varepsilon}\right)\right)$$

où $\varepsilon > 0$ est un réel qui est arbitraire mais fixé, $\varphi(\cdot)$ dénote l'indicatrice d'Euler et $\pi(x)$ désigne le nombre des nombres premiers n'excédant pas x.

Je généralise ce résultat (qui unit des éléments probabilistes et analytiques) et je montre que sous certaines conditions il y a un développement asymptotique des nombres $\mathrm{card}\{n \leqslant x : s_q(g(n)) = k\}$ pour des fonctions $g : \mathbb{N} \to \mathbb{N}$. Les sommes d'exponentielles qui correspondent aux résultats d'autres chapitres, prennent une place très importante. De plus, je traite une question analogue par les entiers de Gauss. En particulier, j'utilise des idées de Gittenberger et Thuswaldner qui ont généralisé la méthode de Bassily et Kátai pour l'anneau des entiers de Gauss et j'obtiens même une amélioration. Enfin, je donne quelques exemples pour montrer l'utilité de ces considérations. Les résultats du Chapitre 6 me permettent de démontrer le théorème suivant :

Théorème. *Soit $c \in \mathbb{R}^+ \setminus \mathbb{N}$. Il existe une constante $q_0(c) \geqslant 2$ telle que pour tout $q \geqslant q_0(c)$ et uniformément en $k \in \mathbb{N}$,*

$$\frac{1}{x}\,\mathrm{card}\{n \leqslant x : s_q(\lfloor n^c \rfloor) = k\}$$
$$= \frac{1}{\sqrt{2\pi\sigma_q^2 c \log_q x}}\left(e^{-\frac{(k-\mu_q c \log_q x)^2}{2\sigma_q^2 c \log_q x}} + O\left(\frac{(\log\log x)^7}{(\log x)^{1/2}}\right)\right).$$

J'obtiens un résultat similaire au développement asymptotique des nombres $\mathrm{card}\{z \in \mathcal{D}_N : s_q(z^2) = k\}$, où $(\mathcal{D}_N)_{n\in\mathbb{N}}$ est un suite de $1/2$-$\mathbb{Z}[i]$ arbitraire. Afin de montrer ce développement, j'emploie un théorème du Chapitre 3 et j'utilise une fois de plus l'inégalité de type van der Corput dans $\mathbb{Z}[i]$.

Finalement, je voudrais résumer les résultats les plus importants de ma thèse, et indiquer les références de mes articles parus, acceptés et soumis. Pour résoudre complètement le problème de Gelfond pour les nombres premiers dans $\mathbb{Z}[i]$, je traite en détail la transformée de Fourier discrète de la somme des chiffres et j'améliore quelques estimations des sommes d'exponentielles connues dans $\mathbb{Z}[i]$ (**Chapitre 2**). Dans les entiers de Gauss je définis des suites de Følner spéciales (des suites de κ-$\mathbb{Z}[i]$) et développe une inégalité de type van der Corput pour traiter la somme des chiffres des carrés (**Chapitre 3 et 7**). L'objet de ces chapitres réside dans mon article [Mor10a] qui est paru dans le *Journal of Number Theory* et mon travail soumis [Mor10b]. De plus, j'étudie des suites de Thue-Morse généralisées dans des groupes topologiques compacts en développant une méthode pour des représentations des groupes par analogie avec la méthode de Mauduit et Rivat (**Chapitre 4**). Ces considérations reposent sur le travail [DM10] que j'ai soumis avec Michael Drmota et qui est accepté dans le *Israel Journal of Mathematics.* Comme application, je calcule des fréquences de lettres dans des sous-suites de carrés des suites automatiques spéciales (inversibles). Pour faciliter l'étude des sous-suites pour les suites automatiques générales, je définis et analyse des fonctions q-multiplicatives généralisées aux matrices complexes (**Chapitre 5**). Quelques résultats de ce chapitre se trouvent dans [DM10]. Enfin, je traite la somme des chiffres de la suite $(\lfloor n^c \rfloor)_{n\in\mathbb{N}}$ pour tout $c \in \mathbb{R}^+ \setminus \mathbb{N}$ en unissant des estimations des sommes d'exponentielles connues et des méthodes de l'analyse harmonique développées récemment (**Chapitre 6 et 7**). Ces résultats se trouvent dans mon travail soumis [Mor10c].

1.3 Introduction in English

Mathematics is the queen
of sciences and arithmetic is
the queen of mathematics.

Carl Friedrich Gauss (1777-1855)

Number theory or, as Gauss called it, arithmetic, is one of the oldest fields of mathematics. Numbers have always fascinated mathematicians all over the world. Euclid showed approximately 2003 years ago that the number of primes is infinite. Since a great part of modern cryptography is based on number theoretic considerations, the study of primes and squares occupy a firm place in our everyday life. First and foremost, the simplicity of its questions and the beauty of its results have made number theory to one of the most flourishing branches in mathematics.

Since the time of Euclid, mathematics has made a remarkable progress. A lot of interesting results could be shown particularly in the theory of numbers. But this theory is also known for various open questions. Goldbach's conjecture (every even integer can be written as a sum of two primes) and the question if there are infinitely many twin primes (a pair of primes whose difference is two) belong to the most famous unsolved problems in mathematics. Not unmentioned should be Riemann's hypothesis, which is a central problem in analytic number theory. An open and unapproachable problem, which is related to this thesis, is the question whether there are infinitely many primes of the form $2^n - 1$ (so-called Mersenne primes) or of the form $2^{2^n} + 1$ (so-called Fermat primes).

These two types of numbers lead me to another branch of number theory. Using the binary system, Mersenne numbers consist only of ones (for example $2^3 - 1 = (111)_2$) and Fermat numbers of exactly one 1 at the beginning and one at the end (for example $2^{2^2} + 1 = (10001)_2$). Although humans already wondered about representations of numbers thousands of years ago, our nowadays commonly used decimal system (including the digit zero) was brought to Europe from India via the Arab World in the Middle Ages. The previously mentioned binary system, which has become so important during the 20th century, was studied in the 17th century for the first time (for a history of number theory and representation systems see [Rib96] and [Knu81, Chapter 4]. There are only few known general results that relate numbers to their digits. Although such problems are often very difficult, there has been some progress in the last decades (see for example [MR10, Chapter 1]). One reason for that may be the fact that in the era of computers, digits have gained in importance.

A crucial role, historically as well as in my thesis, is played by the sum-of-digits function of a number n in base q (denoted by $s_q(n)$). It was studied indirectly in base 2 (where it counts the number of ones in the binary system) by Prouhet in 1851. At the beginning of the 20th century Thue used the sequence $(s_2(n) \bmod 2)_{n \geqslant 0}$, which begins with the digits

$$0\,1\,1\,0\,1\,0\,0\,1\,1\,0\,0\,1\,0\,1\,1\,0\,1\,0\,0\,1\,0\,1\,1\,0\ldots,$$

in order to solve a combinatorial problem. This sequence was rediscovered by several different mathematicians (amongst others by Morse in differential geometric considerations) and is nowadays known as the Thue-Morse sequence (see [Mau01]). First general results concerning the sum-of-digits function where shown in the middle of the 20th century, among others by Bellman, Shapiro, Delange, Kátai and Gelfond. The latter, the Russian mathematician Alexander Osipovich Gelfond (Александр Осипович Гельфонд, 1906 – 1968), contributed substantial achievements to the distribution of the sum-of-digits function. In 1968, in his work *Sur les nombres qui ont des propriétés additives et multiplicatives données* he showed the following result:

Theorem (Gelfond, [Gel68]). *Let q, m and r be positive integers with $q \geqslant 2$ and $(m, q-1) = 1$. Then, for all integers ℓ and a, we have*

$$\#\{1 \leqslant n \leqslant N : n \equiv \ell \bmod r,\, s_q(n) \equiv a \bmod m\} = \frac{N}{mr} + O(N^\lambda),$$

where $\lambda < 1$ is a positive constant only depending on q and m.

At the end of his paper Gelfond stated three questions, which are called Gelfond's sum-of-digits problems. In the following, I will illuminate them and I will state known results and solutions. First, Gelfond conjectured that for all q_1, q_2, m_1 , $m_2 \geqslant 2$ with $(q_1, q_2) = 1$, $(m_1, q_1 - 1) = 1$ and $(m_2, q_2 - 1) = 1$ there exists a positive constant $\lambda < 1$ such that

$$\#\{1 \leqslant n \leqslant N : s_{q_1}(n) \equiv a_1 \bmod m_1 \text{ and } s_{q_2}(n) \equiv a_2 \bmod m_2\} = \frac{N}{m_1 m_2} + O(N^\lambda).$$

A few years later (1972), Bésineau could show an asymptotic result, yet without any estimate of the error term. Almost 30 years later, in 1999, Kim solved Gelfond's first problem. In particular, he proved a more general result for q-additive functions (see [Bés72, Kim99]).

Gelfond's second problem asks for the number of primes p for which the sum-of-digits function $s_q(p)$ lies in a fixed residue class modulo m. The digital nature of the sum-of-digits function and the multiplicative properties of primes, which do not seem to have anything in common, made this question to a highly interesting and until recently unsolved problem. In their work *Sur un problème de Gelfond: la somme des chiffres des nombres premiers*, Mauduit and Rivat showed the following result:

Theorem (Mauduit and Rivat, [MR10]). *Let q and m be integers $\geqslant 2$ and let $d = (q-1, m)$. Then there exists a constant $\sigma_{q,m} > 0$, such that for all $a \in \mathbb{Z}$,*

$$\#\{p \leqslant x : p \textit{ prime and } s_q(p) \equiv a \bmod m\} = \frac{d}{m}\pi(x; a, d) + O(x^{1-\sigma_{q,m}}),$$

where $\pi(x; a, d)$ is equal to the number of primes $p \leqslant x$ with $p \equiv a \bmod d$. Moreover, the sequence $(\alpha s_q(p))$, p prime, is uniformly distributed modulo 1 if and only if $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

The consideration of the sum-of-digits function of polynomial subsequences is the main objective of Gelfond's third question. Let P be a polynomial which satisfies $P(\mathbb{N}) \subseteq \mathbb{N}$. Then, what is the number of positive integers n such that $s_q(P(n)) \equiv a \bmod m$? Interestingly, this seems to be Gelfond's most difficult sum-of-digits problem. This is due to the fact that sparse sequences, i.e., the n-th term of the sequence is much larger than n, make work more difficult. Mauduit and Rivat completely solved Gelfond's third problem in the case of squares ($P(n) = n^2$):

Theorem (Mauduit and Rivat, [MR09])**.** *Let q and m be integers $\geqslant 2$. Let $d = (q-1, m)$ and $Q(a,d) = \# \{0 \leqslant n < d : n^2 \equiv a \bmod d\}$. Then there exists a constant $\sigma_{q,m} > 0$, such that for all $a \in \mathbb{Z}$,*

$$\# \{n \leqslant x : s_q(n^2) \equiv a \bmod m\} = \frac{x}{m} Q(a,d) + O\left(x^{1-\sigma_{q,m}}\right).$$

Moreover, the sequence $(\alpha s_q(n^2))_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 if and only if $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

This thesis, which I will present in the following, deals with general Gelfond problems and related questions. In order to keep the notation in the introduction as simple as possible, at this point, I will sometimes not state my results in full generality. Exact definitions and detailed results can be found in the cited chapters. Since the notation slightly changes in different areas (for instance, $s_q(n)$ denotes the ordinary or the complex sum-of-digits function depending on the underlying ring of integers), I will introduce the specific notation in each chapter. In my thesis I will use the convention that the natural numbers $\mathbb{N}$ include the number 0, since, most notably, zero is a historically (even though a very late excepted) and for the study of the sum-of-digits function crucial digit.

Chapter 2 covers the sum of digits of primes in the ring of Gaussian integers. Inspired by the work of Mauduit and Rivat for the sum of digits of primes in the natural numbers, Drmota, Rivat, and Stoll proved a similar result in the Gaussian integers $\mathbb{Z}[i]$ under certain assumptions (in what follows, $s_q(\,\cdot\,)$ denotes the complex sum-of-digits function):

Theorem (Drmota, Rivat, and Stoll [DRS08])**.** *Let $q = -a \pm i$ be a prime in $\mathbb{Z}[i]$, where $a \geqslant 28$ is a positive integer and $b, g \in \mathbb{Z}$, $g \geqslant 2$. Moreover, set $d = (g, a^2 + 2a + 2)$ and $\delta = (d, 1 \mp i(a+1))$, where the choice of the sign depends on the sign of $q = -a \pm i$. Then there exists $\sigma_{q,g} > 0$ such that*

$$\# \left\{p \in \mathbb{Z}[i] : p \textit{ prime }, |p|^2 \leqslant N, s_q(p) \equiv b \bmod g\right\} = \frac{d}{g}\, \pi_i(N; b, d/\delta) + O\left(N^{1-\sigma_{q,g}}\right),$$

where $\pi_i(N; b, d/\delta)$ denotes the number of Gaussian primes with $|p|^2 \leqslant N$ and $p \equiv b \bmod d/\delta$. Furthermore, the sequence $(\alpha s_q(p))$, running over all Gaussian primes p, is uniformly distributed modulo 1 if and only if $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

In my work I show that one can get rid of the special, somewhat unnatural assumptions of Drmota et al. in order to solve Gelfond's second problem in the Gaussian integers. In particular, I treat all possible bases for which there exists a direct generalization of the sum-of-digits function to $\mathbb{Z}[i]$. Moreover, I analyze primes lying in circular sectors instead of discs. In view of Hecke's prime number theorem, which deals with the number of Gaussian primes lying in circular sectors, the consideration of these regions seem to be a natural choice. In the following we denote the number of Gaussian primes $|p|^2 \leqslant N$ that satisfy $\gamma_1 \leqslant \arg(p) < \gamma_2$ and $p \equiv b \bmod d'$ by $\pi_{\gamma_1,\gamma_2}(N; b, d')$.

Theorem. *Let $q = -a + i$ or $q = -a - i$, $a \geqslant 1$ be a Gaussian integer and $0 \leqslant \gamma_1 < \gamma_2 \leqslant 2\pi$. Furthermore, let $b, g \in \mathbb{Z}$, $g \geqslant 2$ and set $d = (g, a^2 + 2a + 2)$ and $d' = (g, q - 1)$. Then there exists a constant $\sigma_{q,g} > 0$ such that*

$$\#\{|p|^2 \leqslant N : p \text{ prime}, \gamma_1 \leqslant \operatorname{Arg}(p) < \gamma_2,\ s_q(p) \equiv b \bmod g\} = \frac{d}{g}\pi_{\gamma_1,\gamma_2}(N; b, d') + O(N^{1-\sigma_{q,g}}).$$

Let $(p_n)_{n\in\mathbb{N}}$ be a sequence of all Gaussian primes with $\gamma_1 \leqslant \operatorname{Arg}(p_n) < \gamma_2$ ordered such that $|p_{n+1}| \geqslant |p_n|$. Then the sequence $(\alpha s_q(p_n))_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 if and only if $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

The methods used for the proof of this theorem range from a generalized Vaughan identity in $\mathbb{Z}[i]$ to approximation properties of certain functions and a van der Corput-type inequality in the ring of Gaussian integers. Moreover, I develop exponential sum estimates and L^1- and L^∞-norm estimates of the discrete Fourier transform of the complex sum-of-digits function.

The sum-of-digits function of squares is treated in **Chapter 3**. The main focus is placed on the choice of Gaussian integers which I analyze. A sequence $(\mathcal{D}_N)_{N\in\mathbb{N}}$ of subsets of $\mathbb{Z}[i]$ is called a κ-$\mathbb{Z}[i]$ sequence, if for all $N \in \mathbb{N}$ the following four conditions hold true:

(i) $\mathcal{D}_N \subseteq \mathcal{D}_{N+1}$,

(ii) $\mathcal{D}_N \subseteq \{z \in \mathbb{Z}[i] : \max(|\Re(z)|, |\Im(z)|) \leqslant \sqrt{N}\}$,

(iii) there exists a constant $c > 0$, such that $cN \leqslant \#\mathcal{D}_N$, and

(iv) $\#\{\mathcal{D}_N \triangle (r + \mathcal{D}_N)\} \ll |r|N^{1-\kappa}$ for all $r \in \mathbb{Z}[i]$.

These sequences are special Følner sequences where the rate of convergence comes into play. The simplest examples of κ-$\mathbb{Z}[i]$ sequences with $\kappa = 1/2$ are Gaussian integers lying in squares with side length $\sqrt{N}$ or discs with radius $\sqrt{N}$ (see Figure 1.3). Interestingly, one can show that every sequence of convex sets satisfying conditions (i), (ii) and (iii) is a $1/2$-$\mathbb{Z}[i]$ sequence. The main result of this chapter is the following, where $Q(b, d)$ denotes the number of Gaussian integers z in a complete residue system modulo d with $z^2 \equiv b \bmod d$:

Theorem. *Let $q = -a + i$ or $q = -a - i$ and $a \geqslant 1$ an integers, such that every prime divisor $p \mid q$ satisfies $|p| \geqslant \sqrt{689}$. Moreover, let $b, g \in \mathbb{Z}$, $g \geqslant 2$ and $(\mathcal{D}_N)_{N\in\mathbb{N}}$ a κ-$\mathbb{Z}[i]$ sequence with $0 < \kappa \leqslant 1/2$. Set $d = (g, q-1)$. Then there exists a constant $\sigma_{q,g,\kappa} > 0$ such that*

$$\#\left\{z \in \mathcal{D}_N : s_q(z^2) \equiv b \bmod g\right\} = \frac{\#\mathcal{D}_N}{g} Q(b,d) + O\left(N^{1-\sigma_{q,g,\kappa}}\right).$$

Furthermore, if $(z_n)_{n\in\mathbb{N}}$ is a sequence of all Gaussian integers ordered such that $|z_{n+1}| \geqslant |z_n|$, then the sequence $(\alpha s_q(z_n^2))_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 if and only if α is irrational.

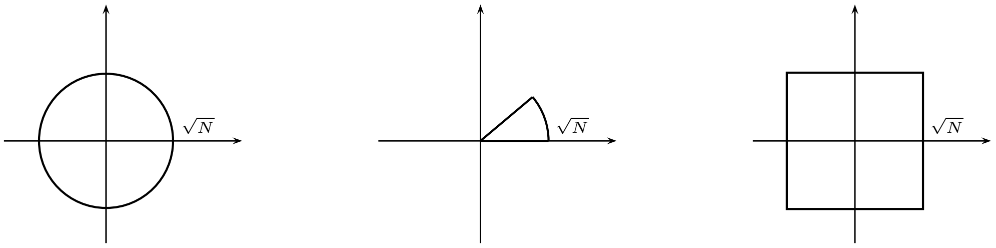


Figure 1.3: κ-$\mathbb{Z}[i]$ sequences with $\kappa = 1/2$

Similarly to the case of prime numbers, I use a van der Corput-type inequality in the Gaussian integers. However, the remarkable property of this inequality is, that it allows the study of the complex sum-of-digits function in κ-$\mathbb{Z}[i]$ sequences. Furthermore, I obtain estimates of Gauss sums in $\mathbb{Z}[i]$. Together with the fact that addition in the ring of Gaussian integers can be realized using an automaton, in the end these sums make it possible to employ the discrete Fourier transform of the sum-of-digits function. Exponential sum estimates (results of Gittenberger and Thuswaldner are improved) as well as the Fourier theoretic results obtained in Chapter 2 complete the proof.

Chapter 4 contains a generalization of the Thue-Morse sequence and the ordinary sum-of-digits function to group-valued functions. Let $q \geqslant 2$, H be a compact group that satisfies the Hausdorff separation axiom, and $g_0, g_1, \ldots, g_{q-1} \in H$ with $g_0 = e$ the identity element. Furthermore, let $G \leqslant H$ be the closure of the subgroup generated by $g_0, g_1, \ldots, g_{q-1}$. Suppose that $\varepsilon_{0(n)}, \ldots, \varepsilon_{\ell-1(n)}$ are the digits of the integer n in base q. The sequence

$$T(n) = g_{\varepsilon_0(n)} g_{\varepsilon_1(n)} \cdots g_{\varepsilon_{\ell-1(n)}}$$

$n \geqslant 0$, is called a generalized Thue-Morse sequence. It is a completely q-multiplicative G-valued sequence and if $G = \mathbb{Z}/2\mathbb{Z}$ (with $+$ as the group operation), $q = 2$, and $g_0 = 0$, $g_1 = 1$, then $T(n)$ corresponds to the classical Thue-Morse sequence. The sequence $(\alpha s_q(n))_{n\in\mathbb{N}}$ with a real number α can be realized on the torus.

Theorem. *Let $(T(n))_{n\in\mathbb{N}}$ be a generalized Thue-Morse sequence. Moreover, let a and b be natural numbers with $a \geqslant 1$. Then there exist absolutely continuous measures ν_1 and ν_2, such that $T(an+b)_{n\in\mathbb{N}}$ is ν_1-distributed in G and $T(n^2)_{n\in\mathbb{N}}$ is ν_2-distributed in G.*

The theory of compact group representations allows to determine these measures in a unique manner. Moreover, it is possible to state a criterion which decides whether the considered sequences are uniformly distributed with respect to the Haar measure. Irreducible and unitary representations occupy a firm place in the reasoning of the given proof, which is based on a proper generalization of the Fourier-theoretic method of Mauduit and Rivat.

Chapter 5 treats automatic sequences and frequencies of letters in subsequences. First, I show an application of the result concerning the generalized Thue-Morse sequence $(T(n))_{n\in\mathbb{N}}$. As it turns out, there is a close relation between $(T(n))_{n\in\mathbb{N}}$ and so-called invertible automatic sequences. In the case of finite groups, the distribution properties of $(T(n^2))_{n\in\mathbb{N}}$ imply the following result:

Theorem. *Let $q \geqslant 2$ and $(u_n)_{n\in\mathbb{N}}$ an invertible q-automatic sequence over an alphabet Δ. Moreover, let $a \in \Delta$. Then there exists the frequency of a in the sequence $(u_{n^2})_{n\in\mathbb{N}}$.*

Unfortunately, Thue-Morse sequences do not imply results for automatic sequences without any constraints. However, with the help of a method worked out by Mauduit and Rivat, the automatic structure of the given sequences makes it possible to establish a connection between the frequency of letters in general q-automatic sequences $(u_n)_{n\in\mathbb{N}}$ and its subsequences $(u_{\lfloor n^c\rfloor})_{n\in\mathbb{N}}$ for $1 < c < 7/5$:

Theorem. *Let $c \in (1, 7/5)$, $q \geqslant 2$ and $(u_n)_{n\geqslant 0}$ a q-automatic sequence. Furthermore, let a be a letter occurring in $(u_n)_{n\geqslant 0}$. Then the logarithmic frequency of a in $(u_{\lfloor n^c\rfloor})_{n\geqslant 0}$ exists and it is the same as the logarithmic frequency of a in $(u_n)_{n\geqslant 0}$. Moreover, the frequency of a in $(u_n)_{n\geqslant 0}$ exists if and only if the frequency of a in $(u_{\lfloor n^c\rfloor})_{n\geqslant 0}$ exists.*

In order to prove this theorem I generalize a result of Mauduit and Rivat [MR05, Theorem 2] to q-multiplicative functions in the ring of complex square matrices. The universal applicability of the double large sieve of Bombieri and Iwaniec is one of the key points in the reasoning of the proof.

Chapter 6 deals with the sum-of-digits function in the natural numbers of the sequence $(\lfloor n^c\rfloor)_{n\in\mathbb{N}}$, where c is a positive real number different from an integer. The formulation of this problem is linked to Gelfond's third question, where polynomials are replaced by similarly increasing functions. It can be understood as an intermediate case between polynomials of different degree. Mauduit and Rivat [MR95, MR05] studied q-multiplicative properties of the sequence $(\lfloor n^c\rfloor)_{n\in\mathbb{N}}$ for $c \in (1, 7/5)$. In particular, they show the following result:

Theorem (Mauduit and Rivat, [MR05]). *Let $c \in (1, 7/5)$ and $q \geqslant 2$. Then, for all a, $m \in \mathbb{N}$ with $m \geqslant 1$,*

$$\lim_{x\to\infty} \frac{1}{x} \# \{n \leqslant x : s_q(\lfloor n^c \rfloor) \equiv a \bmod m\} = \frac{1}{m}.$$

Furthermore, the sequence $(\alpha s_q(\lfloor n^c \rfloor))_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 if and only if $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

It seems to be difficult to show such a general result for big values of c. However, using Fourier theoretic methods, I can show that for each positive real number c that is different from an integer the sequence $s_q(\lfloor n^c \rfloor)$ is well distributed for sufficiently large bases q.

Theorem. *Let $c \in \mathbb{R}^+ \setminus \mathbb{N}$ and $a, m \in \mathbb{N}$ with $m \geqslant 1$. Then there exists a constant $q_0(c) \geqslant 2$, such that for all $q \geqslant q_0(c)$*

$$\# \{n \leqslant x : s_q(\lfloor n^c \rfloor) \equiv a \bmod m\} = \frac{x}{m} + O\left(x^{1-\sigma_{q,m,c}}\right)$$

with a constant $\sigma_{q,m,c} > 0$. Furthermore, the sequence $(\alpha s_q(\lfloor n^c \rfloor))_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 for all $q \geqslant q_0(c)$ if and only if $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Mauduit [Mau01] conjectured that for almost all $c > 1$ and for all $q, m \geqslant 2$,

$$\lim_{x\to\infty} \frac{1}{x} \# \{n \leqslant x : s_q(\lfloor n^c \rfloor) \equiv a \bmod m\} = \frac{1}{m}. \tag{1.3}$$

My result does not solve this conjecture entirely, but it leads me to conjecture that (1.3) is valid if and only if $c \notin \mathbb{N}$, $c > 1$. (It is easy to see that equation (1.3) cannot hold true for integers $c > 1$.) In order to show the cited result for as many bases as possible, I use two different methods. On the one hand, I adapt a method used by Deshouillers (see [Des72] and [Des73a, Des73b]), which works for all real numbers $c > 1$ different from an integer. On the other hand, in the range $(1, 19/11)$ I modify a method of Mauduit and Rivat [MR05] and I obtain a stronger result in this case. In addition to the mentioned Fourier transform, I use in both methods classical exponential sum estimates à la van der Corput and Vinogradov as well as approximation properties of the Beurling-Selberg function.

In **Chapter 7** I consider local distribution results of the ordinary and the complex sum-of-digits function. Drmota, Mauduit, and Rivat improved a method developed by Bassily and Kátai in order to show the following result (μ_q and σ_q are defined by $\mu_q = (q-1)/2$ and $\sigma_q^2 = (q^2-1)/12$):

Theorem (Drmota, Mauduit, and Rivat, [DMR09]). *Let $q \geqslant 2$. We then have uniformly in $k > 0$ with $(k, q-1) = 1$*

$$\begin{aligned}&\# \{p \leqslant x : p \textit{ prim}, \ s_q(p) = k\} \\ &\qquad = \frac{q-1}{\varphi(q-1)} \frac{\pi(x)}{\sqrt{2\pi\sigma_q^2 \log_q x}} \left(e^{-\frac{(k-\mu_q \log_q x)^2}{2\sigma_q^2 \log_q x}} + O\left((\log x)^{-1/2+\varepsilon}\right)\right),\end{aligned}$$

where $\varepsilon > 0$ is arbitrary but fixed, $\varphi(\,\cdot\,)$ denotes Euler's totient function and $\pi(x)$ is equal to the number of primes less than or equal to x.

I generalize the work of Drmota, Mauduit, and Rivat and show that, under certain assumptions on a function $g : \mathbb{N} \to \mathbb{N}$, I get an asymptotic expansion of the numbers $\#\{n \leqslant x : s_q(g(n)) = k\}$. Furthermore, I discuss analog problems in the ring of Gaussian integers. Developing ideas used by Gittenberger and Thuswaldner, I obtain similar results in the complex setting. Finally, I give some examples in order to emphasize the utility of the general considerations. In particular, exponential sum estimates proved in Chapter 6 permit to show the following result:

Theorem. *Let $c \in \mathbb{R}^+ \setminus \mathbb{N}$. Then there exists a constant $q_0(c) \geqslant 2$, such that for all $q \geqslant q_0(c)$ and uniformly in $k \in \mathbb{N}$*

$$\frac{1}{x}\#\{n \leqslant x : s_q(\lfloor n^c \rfloor) = k\}$$
$$= \frac{x}{\sqrt{2\pi\sigma_q^2 c \log_q x}} \left(e^{-\frac{(k-\mu_q c \log_q x)^2}{2\sigma_q^2 c \log_q x}} + O\left(\frac{(\log\log x)^7}{(\log x)^{1/2}} \right) \right).$$

Moreover, results in Chapter 3 together with a van der Corput-type inequality in $\mathbb{Z}[i]$ imply an asymptotic expansion of the numbers $\#\{z \in \mathcal{D}_N : s_q(z^2) = k\}$ for arbitrary κ-$\mathbb{Z}[i]$ sequences $(\mathcal{D}_N)_{n\in\mathbb{N}}$ with $\kappa = 1/2$.

Finally, I would like to summarize the most important results of my thesis and give references to my submitted, accepted, and published papers. In order to completely solve Gelfond's problem for prime numbers in the Gaussian integers, I treat the discrete Fourier transform of the complex sum-of-digits function in detail and improve some known exponential sum estimates in $\mathbb{Z}[i]$ (**Chapter 2**). Furthermore, I introduce special Følner-sequences (so-called κ-$\mathbb{Z}[i]$ sequences) in the Gaussian integers and prove a van der Corput-type inequality as to show distribution results for the sum-of-digits function of squares (**Chapters 3 and 7**). The content of these chapters is based on my work [Mor10a], which was published in the *Journal of Number Theory*, and my submitted paper [Mor10b]. Next, I study generalized Thue-Morse sequences in compact groups by developing a method for group representations analogous to the Fourier based method of Mauduit and Rivat (**Chapter 4**). These considerations are based on the paper [DM10], which is joint work with Michael Drmota and which was accepted for publication in the *Israel Journal of Mathematics*. As an application, I consider the frequency of letters in the subsequence of squares of invertible q-automatic sequences. In order to obtain results for subsequences of general automatic sequences, I introduce and analyze matrix-valued q-multiplicative functions (**Chapter 5**). Some parts of these results can be found in [DM10]. Finally, I show that the sum-of-digits function of the sequence $(\lfloor n^c \rfloor)_{n\in\mathbb{N}}$ for $c \in \mathbb{R}^+ \setminus \mathbb{N}$ is uniformly distributed by combining well-known exponential sum estimates and recently developed methods in harmonic analysis (**Chapters 6 and 7**). These results are based on my submitted paper [Mor10c].

Chapter 2

The sum of digits of Gaussian primes

The mathematician's patterns, like the painter's
or the poet's must be beautiful; the ideas, like the colors
or the words must fit together in a harmonious way.
Beauty is the first test: there is no
permanent place in this world for ugly mathematics.

Godfrey Harold Hardy (1877 - 1947)

In this chapter we consider the sum of digits of primes in the ring of Gaussian integers. Let $q = -a \pm i$ and denote by s_q the complex sum-of-digits function with respect to base q (see Apendix A.1). We show that the sequence $(\alpha s_q(p))$ running over all Gaussian primes lying in a circular sector is uniformly distributed modulo 1 if and only if α is irrational. Moreover, we prove that the sum-of-digits function of primes is well distributed in arithmetic progressions. In particular, we determine the order of magnitude of the number of Gaussian primes lying in a disc whose sum-of-digits evaluation lies in some fixed residue class mod m. This work generalizes a theorem of Mauduit and Rivat that was the solution of a long standing conjecture by Gelfond concerning the usual q-ary sum-of-digits function. It also improves a result of Drmota, Rivat, and Stoll, who could only deal with sufficiently large prime bases $q = -a \pm i$ and the full disc instead of circular sectors.

2.1 Introduction and main results

Before stating our main result, we recall the historical development of Gelfond's sum-of-digits problem for primes. Gelfond [Gel68] remarked in 1968, that it would be interesting to find the number of primes p less than or equal to x, such that $s_q(p) \equiv b \bmod m$, where $s_q(n)$ denotes the q-ary sum-of-digits function in $\mathbb{N}$. Based on a Fourier theoretic method, Mauduit and Rivat [MR10] could recently show that the sum of digits of primes is well distributed in arithmetic progressions: Let m be an integer $\geqslant 2$ and set[1] $d = (m, q-1)$. Then there exists a constant $\sigma_{q,m} > 0$, such that for every $b \in \mathbb{Z}$,

$$\# \{p \leqslant x : p \text{ prime and } s_q(p) \equiv b \bmod m\} = \frac{d}{m}\pi(x; d, b) + O_{q,m}(x^{1-\sigma_{q,m}}),$$

where $\pi(x; d, b)$ denotes the number of primes $p \leqslant x$, such that $p \equiv b \bmod d$. Moreover, they proved that the sequence $(\alpha s_q(p))$, running over all primes, is uniformly distributed modulo 1 if and only if α is irrational. Here and in what follows, the symbol $f = O_\omega(g)$ means that $|f| \leqslant cg$, where the constant c may depend on ω.

Drmota, Rivat, and Stoll [DRS08] considered the same problem in the ring of Gaussian integers. If $q = -a \pm i$ (choose a sign), where $a \in \mathbb{Z}^+$, then every $z \in \mathbb{Z}[i]$ has a unique finite representation of the form

$$z = \sum_{j \geqslant 0} \varepsilon_j(z) q^j, \quad \varepsilon_j(z) \in \{0, 1, \dots, |q|^2 - 1\},$$

with $\varepsilon_j(z) = 0$, for j greater than some constant $j_0(z)$. The number $\varepsilon_j(z)$ is called the j-th digit of the number z in the base-q representation system. The Gaussian integers $q = -a \pm i$, $a \geqslant 1$, are the only possible bases such that the digits are given by $\{0, 1, \dots, |q|^2 - 1\}$ (see Theorem A.2). The (complex) sum-of-digits function s_q is then defined by $z \mapsto \sum_{j \geqslant 0} \varepsilon_j(z)$ (see Appendix A.1 for further information on the base-q representation system in $\mathbb{Z}[i]$ and the sum-of-digits function s_q).

Theorem DRS (Drmota, Rivat, and Stoll [DRS08]). *Let $q = -a \pm i$ be a prime in $\mathbb{Z}[i]$ with $a \geqslant 28$ a positive integer and let $b, g \in \mathbb{Z}$, $g \geqslant 2$. Moreover, set $d = (g, a^2 + 2a + 2)$ and $\delta = (d, 1 \mp i(a+1))$, where the choice of the sign depends on the sign of $q = -a \pm i$. Then there exists $\sigma_{q,g} > 0$ such that*

$$\# \left\{p \in \mathbb{Z}[i] : p \textit{ prime }, |p|^2 \leqslant N, s_q(p) \equiv b \bmod g\right\} = \frac{d}{g}\, \pi_i(N; b, d/\delta) + O_{q,g}\left(N^{1-\sigma_{q,g}}\right),$$

where $\pi_i(N; b, d/\delta)$ denotes the number of Gaussian primes with $|p|^2 \leqslant N$ and $p \equiv b \bmod d/\delta$. Furthermore, the sequence $(\alpha s_q(p))$, running over all Gaussian primes p, is uniformly distributed modulo 1 if and only if α is irrational.

The assumptions on the base q ($q = -a \pm i$ is prime and $a \geqslant 28$) in Theorem DRS arise from the method of the proof given in [DRS08]. In the end, the similarity to the circle problem makes it impossible for Drmota et al. to treat composite bases and small prime bases. Note, that it is not known if there exist infinitely many Gaussian primes of the form $q = -a \pm i$ (since it is not known if there exist infinitely many primes of the form $a^2 + 1$). The restriction in Theorem DRS to primes lying in a disc $\{z \in \mathbb{Z}[i] : |z|^2 \leqslant N\}$ is also a drawback of the method used for the proof of the

[1] We use throughout this thesis following standard notation: If $a, b \in \mathbb{N}$, then (a, b) denotes the unique greatest common divisor of a and b. If $a, b \in \mathbb{Z}[i]$, then the same symbol denotes an arbitrary but fixed greatest common divisor in $\mathbb{Z}[i]$ of a and b.

theorem. However, as it is remarked in [DRS08] (see page 322), it would be natural to consider the distribution of the sum-of-digits function of Gaussian primes lying in angular regions. In view of Hecke's prime number theorem (which deals with the number of Gaussian primes lying in circular sectors, see below), the consideration of these regions seem to be a natural choice.

In what follows, we show that we can get rid of these special assumptions in Theorem DRS. In particular, we prove that Theorem DRS holds true for all possible bases $q = -a \pm i$, $a \geqslant 1$ and for Gaussian primes lying in an circular sector of the form[2]

$$\{n \in \mathbb{Z}[i] : |n|^2 \leqslant N, \gamma_1 \leqslant \operatorname{Arg}(n) < \gamma_2\},$$

where $0 \leqslant \gamma_1 < \gamma_2 \leqslant 1$ (note, that $\gamma_1 = 0$ and $\gamma_2 = 1$ corresponds to the disc). Most notably, improved Fourier transform and exponential sum estimates as well as special approximation properties of certain functions permit to obtain a result in this generality. The main focus is placed on the treatment of the exponential sum

$$\sum_{\substack{|n|^2 \leqslant N \\ \gamma_1 \leqslant \operatorname{Arg}(n) < \gamma_2}} \Lambda_i(n)\, \mathrm{e}(\alpha s_q(n)),$$

where $0 \leqslant \gamma_1 < \gamma_2 \leqslant 1$, $\Lambda_i(n)$ denotes the complex von Mangoldt function

$$\Lambda_i(n) = \begin{cases} \log |p|, & \text{if } n = \varepsilon p^\nu, \text{ for some unit } \varepsilon, \text{ prime } p \text{ and integer } \nu > 0, \\ 0, & \text{otherwise,} \end{cases}$$

and $\mathrm{e}(x)$ is defined by $e^{2\pi i x}$.

Theorem 2.1. *Let $q = -a \pm i$, $a \geqslant 1$ and $0 \leqslant \gamma_1 < \gamma_2 \leqslant 1$. Then for any $\alpha \in \mathbb{R}$ with $(a^2 + 2a + 2)\alpha \notin \mathbb{Z}$ there exists a constant $\sigma_q(\alpha) > 0$ such that*[3]

$$\sum_{\substack{|n|^2 \leqslant N \\ \gamma_1 \leqslant \operatorname{Arg}(n) < \gamma_2}} \Lambda_i(n)\, \mathrm{e}(\alpha s_q(n)) \ll_{q,\alpha} N^{1-\sigma_q(\alpha)}.$$

The following two corollaries are a direct consequence of Theorem 2.1. For the proof of Corollary 2.2, we use in addition Hecke's prime number theorem (see [HST91, page 126] or [GL66, Chapter 4]), which states that the number of Gaussian primes $|p|^2 \leqslant x$ with $\gamma_1 \leqslant \operatorname{Arg}(p) < \gamma_2$ is asymptotically given by $4(\gamma_2 - \gamma_1)x/\log x$.

Corollary 2.2. *Let $q = -a \pm i$, $a \geqslant 1$ and $0 \leqslant \gamma_1 < \gamma_2 \leqslant 1$. Furthermore, let $(p_n)_{n\in\mathbb{N}}$ be a sequence of all Gaussian primes with $\gamma_1 \leqslant \operatorname{Arg}(p_n) < \gamma_2$ ordered such that $|p_{n+1}| \geqslant |p_n|$. Then the sequence $(\alpha s_q(p_n))_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 if and only if α is irrational.*

[2] If $z \neq 0$, we denote by $\operatorname{Arg}(z)$ the unique real number $0 \leqslant \kappa < 1$, such that $z = |z|e^{2\pi i\kappa}$.

[3] The symbol $f \ll g$ means that $f = O(g)$ and $f \ll_\omega g$ means that $f = O_\omega(g)$. If not otherwise stated, this convention is used throughout the whole thesis.

Corollary 2.3. *Let $q = -a \pm i$, $a \geqslant 1$ and $0 \leqslant \gamma_1 < \gamma_2 \leqslant 1$. Furthermore, let b, $g \in \mathbb{Z}$, $g \geqslant 2$ and set $d = (g, a^2 + 2a + 2)$. Then there exists a constant $\sigma_{q,g} > 0$ such that*

$$\#\{|p|^2 \leqslant N : p \text{ prime}, \gamma_1 \leqslant \mathrm{Arg}(p) < \gamma_2, s_q(p) \equiv b \bmod g\}$$
$$= \frac{d}{g}\pi_{\gamma_1,\gamma_2}(N; b, (g, q-1)) + O_{q,g}(N^{1-\sigma_{q,g}}),$$

where $\pi_{\gamma_1,\gamma_2}(N; b, (g, q-1))$ denotes the number of Gaussian primes $|p|^2 \leqslant N$ with $\gamma_1 \leqslant \mathrm{Arg}(p) < \gamma_2$ and $p \equiv b \bmod (g, q-1)$.

This chapter is organized as follows. The next Section deals with L^1-norm and L^∞-norm estimates of the discrete Fourier transform of the sum-of-digits function in $\mathbb{Z}[i]$. The starting point of the proof of Theorem 2.1 is Vaughan's identity in $\mathbb{Z}[i]$, which allows to split up the problem in estimating so-called type I sums and type II sums (see Section 2.3.1). We give a new proof of this identity in order to be able to treat circular sectors (see also Remark 2.11). Since type I sums for discs are already treated in [DRS08] (and since the reasoning for circular sectors is very similar), we give in Section 2.3.2 just a short outline of the proof of the type I sum estimate. The main task of this chapter is the proof of the type II sum estimate for general bases and circular sectors (see Section 2.3.3).

We start with "removing" the multiplicative constraints of the indices of summation as well as the condition coming from the fact that we are dealing with circular sectors (Lemma 2.14). In order to do this, we use special approximation properties related to Perron's summation formula. Using approximation properties of the Beurling-Selberg function together with the Koksma-Hlawka inequality, we give a second proof of Lemma 2.14 (for the sake of understanding). We continue with showing a van der Corput-type inequality and a "carry" lemma which allow us to truncate the sum-of-digits function. (Comparable results can be found in [DRS08]. Since there are some inconsistencies in the proofs, we give new ones, see also Remarks 2.16 and 2.18.) Next, we can avoid problems similar to the circle problem and Fourier theoretic methods lead us to expressions that are related to linear exponential sums, see Remark 2.19 and Lemmas 2.20 and 2.21. These two lemmas are of considerable importance for the proof of the main theorem. In the end, together with the Fourier transform estimates, they make it possible, that we can show the desired result in its full generality. In this part of the proof we use also ideas developed by Martin, Mauduit, and Rivat [MMR] in order to simplify the calculations. In Section 2.3.4 we finish the proof of Theorem 2.1. Section 2.4 finally contains the proofs of Corollary 2.2 and Corollary 2.3.

2.2 Fourier analysis of F_λ

In what follows we denote by R_m a complete residue system modulo m ($m \in \mathbb{Z}[i]$). Moreover, we set $Q = |q|^2 = a^2 + 1$ and write $\|x\|$ for the distance from $x \in \mathbb{R}$ to its nearest integer. Recall, that we use $\mathrm{e}(x)$ for the exponential function $e^{2\pi i x}$. Let

$\mathcal{F}_\lambda = \{\sum_{j=0}^{\lambda-1} \varepsilon_j q^j : \varepsilon_j \in \{0, 1, \dots |q|^2 - 1\}\}$ be the finite (non-scaled) approximation of the fundamental region of the number system. Note, that it is also a complete system of residues $\bmod q^\lambda$. The discrete Fourier transform $F_\lambda(\,\cdot\,, \alpha)$ of the function $u \mapsto \mathrm{e}(\alpha s_q(u))$ is defined for all $h \in \mathbb{Z}[i]$ by

$$F_\lambda(h, \alpha) = Q^{-\lambda} \sum_{u \in \mathcal{F}_\lambda} \mathrm{e}\left(\alpha s_q(u) - \frac{1}{2}\operatorname{tr}(huq^{-\lambda})\right),$$

where $\operatorname{tr}(z) = z + \bar{z}$ denotes the trace of z. Since $F_\lambda(\,\cdot\,, \alpha)$ is periodic with period q^λ and $|F_0(h, \alpha)| = 1$, we have

$$|F_\lambda(h, \alpha)| = Q^{-\lambda} \prod_{j=1}^{\lambda} \varphi_Q\left(\alpha - \frac{1}{2}\operatorname{tr}(hq^{-j})\right), \tag{2.1}$$

where the function $\varphi_k(t)$ is defined for all $k \geqslant 2$ by

$$\varphi_k(t) = \begin{cases} |\sin(\pi k t)|/|\sin(\pi t)|, & t \in \mathbb{R}/\mathbb{Z}, \\ k, & \text{otherwise.} \end{cases}$$

In order to find an upper bound of $|F_\lambda(h, \alpha)|$, we show the following lemma. It is based on a thorough analysis of the proof of [MR09, Lemma 6] (with three adjacent terms instead of two) and generalizes the cited lemma to the Gaussian integers.

Lemma 2.4. *Let $a \geqslant 1$ and $q = -a \pm i$. We have for $\alpha \in \mathbb{R}$ that*

$$\begin{aligned}\max_{z \in \mathbb{C}} \varphi_Q\left(\alpha - \frac{1}{2}\operatorname{tr}(z)\right) \varphi_Q\left(\alpha - \frac{1}{2}\operatorname{tr}(qz)\right) \varphi_Q\left(\alpha - \frac{1}{2}\operatorname{tr}(q^2 z)\right) \\ \leqslant Q^2 \varphi_Q\left(\frac{\|(a^2 + 2a + 2)\alpha\|}{a^2 + 2a + 2}\right).\end{aligned}$$

Proof. For readability, we set $\delta = \frac{\|(a^2+2a+2)\alpha\|}{a^2+2a+2} \leqslant \frac{1}{2Q}$, $u = (a^2 + 2a + 2)\alpha$, $v = (\alpha - (1/2)\operatorname{tr}(qz))$ and $w = \left(\alpha - (1/2)\operatorname{tr}(q^2 z)\right)$. We have to show that at least one of the factors is bounded by $\varphi_Q(\delta)$. Taking into account part (i) of Lemma A.6 it suffices to study the case $\|v\| < \delta$ and $\|w\| < \delta$. Note, that there exist integers u_1, v_1, w_1 and $\varepsilon_u, \varepsilon_v, \varepsilon_w \in \{\pm 1\}$ such that $u = u_1 + \varepsilon_u \|u\|$, $v = v_1 + \varepsilon_v \|v\|$ and $w = w_1 + \varepsilon_w \|w\|$. We use the identity

$$(1 + a^2)\left(\alpha - \frac{1}{2}\operatorname{tr}(z)\right) = u - 2av - w,$$

to obtain (note that $Q = a^2 + 1$)

$$\left\| Q\left(\alpha - \frac{1}{2}\operatorname{tr}(z)\right)\right\| = \|\varepsilon_u \|u\| - 2a\varepsilon_v \|v\| - \varepsilon_w \|w\|\|.$$

Next we claim that

$$\left\| Q\left(\alpha - \frac{1}{2}\operatorname{tr}(z)\right)\right\| \geqslant Q\delta.$$

In order to prove this, it suffices to show that

$$\delta(a^2+1) \leqslant |\varepsilon_u \|u\| - 2a\varepsilon_v \|v\| - \varepsilon_w \|w\|| \leqslant 1-\delta(a^2+1).$$

Using the triangle inequality, we obtain

$$\begin{aligned}|\varepsilon_u \|u\| - 2a\varepsilon_v \|v\| - \varepsilon_w \|w\|| &\leqslant \delta(a^2+2a+2)+2a\delta+\delta \\ &= 2\delta(a^2+2a+2)-\delta(a^2+1) \leqslant 1-\delta(a^2+1).\end{aligned}$$

The last inequality follows from the fact that $\delta(a^2+2a+2) = \left\|(a^2+2a+2)\alpha\right\|$. On the other hand, the inverse triangle inequality gives us

$$|\varepsilon_u \|u\| - 2a\varepsilon_v \|v\| - \varepsilon_w \|w\|| \geqslant \delta(a^2+2a+2) - 2a\delta - \delta = \delta(a^2+1).$$

We finally obtain

$$\varphi_Q\left(\alpha - \frac{1}{2}\operatorname{tr}(z)\right) = \varphi_Q\left(\frac{Q\left(\alpha-\frac{1}{2}\operatorname{tr}(z)\right)}{Q}\right) \leqslant \varphi_Q\left(\frac{\left\|Q\left(\alpha-\frac{1}{2}\operatorname{tr}(z)\right)\right\|}{Q}\right) \leqslant \varphi_Q(\delta).$$

To show the last two inequalities we used parts (v) and (i) of Lemma A.6. □

Lemma 2.5. *Let* $\tilde{c}_Q = \frac{\pi^2}{18\log Q}\left(\frac{Q^2-1}{(a^2+2a+2)^2}\right)$. *Then we have for* $\lambda \geqslant 3$,

$$|F_\lambda(h,\alpha)| \leqslant e^{\pi^2/36} Q^{-\tilde{c}_Q \left\|(a^2+2a+2)\alpha\right\|^2 \lambda}.$$

Proof. First we note, that we have by the previous lemma and by part (ii) of Lemma A.6

$$\begin{aligned}&\max_{z\in\mathbb{C}} \varphi_Q\left(\alpha-\frac{1}{2}\operatorname{tr}(z)\right)\varphi_Q\left(\alpha-\frac{1}{2}\operatorname{tr}(qz)\right)\varphi_Q\left(\alpha-\frac{1}{2}\operatorname{tr}(q^2z)\right) \\ &\qquad \leqslant Q^3\exp\left(-\frac{\pi^2(Q^2-1)}{6(a^2+2a+2)^2}\left\|(a^2+2a+2)\alpha\right\|^2\right).\end{aligned}$$

Let $\lambda \geqslant 3$. Then we obtain

$$\begin{aligned}|F_\lambda(h,\alpha)| &\leqslant \prod_{j=1}^{\lfloor\lambda/3\rfloor} \frac{\varphi_Q\left(\alpha-\frac{1}{2}\operatorname{tr}\left(\frac{h}{q^{3j}}\right)\right)\varphi_Q\left(\alpha-\frac{1}{2}\operatorname{tr}\left(\frac{h}{q^{3j-1}}\right)\right)\varphi_Q\left(\alpha-\frac{1}{2}\operatorname{tr}\left(\frac{h}{q^{3j-2}}\right)\right)}{Q^3} \\ &\leqslant \exp\left(-\frac{\pi^2(Q^2-1)}{6(a^2+2a+2)^2}\lfloor\lambda/3\rfloor\left\|(a^2+2a+2)\alpha\right\|^2\right)\end{aligned}$$

Since $\lfloor\lambda/3\rfloor \geqslant (\lambda-2)/3$, this implies the desired result. □

Remark 2.6. A slightly weaker uniform upper bound of $|F_\lambda(h,\alpha)|$ can be obtained by improving [DRS08, Corollary 6.4]: Let $c_Q = \frac{\pi^2}{27\log Q}((Q^2-1)/Q^4)$. Then we have for $\lambda \geqslant 3$,

$$|F_\lambda(h,\alpha)| \leqslant e^{\pi^2/(54Q^2)} Q^{-c_Q\left\|(a^2+2a+2)\alpha\right\|^2\lambda}. \tag{2.2}$$

Since $2/(3Q) \leqslant (6/(\pi^2(Q^2-1)))^{1/2}$, we obtain from point (i) and (ii) of Lemma A.6 that we have $\varphi_Q(t) \leqslant Q\exp(-(Q^2-1)\pi^2 \|t\|^2/6)$ if $\|t\| \leqslant 2/(3Q)$ and that $\varphi_Q(t) \leqslant \varphi_Q(2/(3Q)) \leqslant Q\exp(-(Q^2-1)\pi^2(2/(3Q))^2/6\|t\|^2)$ if $\|t\| > 2/(3Q)$. We obtain that for all $t \in \mathbb{R}$ the following estimate holds:

$$\varphi_Q(t) \leqslant Q\exp\left(-\frac{2\pi^2(Q^2-1)}{27Q^2}\|t\|^2\right).$$

Carrying out exactly the same steps as in the proof of [DRS08, Corollary 6.4], we get

$$\begin{aligned}|F_\lambda(h,\alpha)| &\leqslant \exp\left(-\frac{\lambda-2}{2Q^2}\cdot\frac{2\pi^2(Q^2-1)}{27Q^2}\left\|(a^2+2a+2)\alpha\right\|^2\right)\\ &\leqslant e^{\pi^2/(54Q^2)}Q^{-c_Q\left\|(a^2+2a+2)\alpha\right\|^2\lambda}.\end{aligned}$$

This finally shows (2.2).

The next lemma of this section gives an L^1-type upper bound of F_λ. It is an improvement of [DRS08, Lemma 6.6] and enables us to consider composite bases in Theorem 2.1 instead of just prime bases. In view of an applications of this lemma in Chapter 3, we prove it in a more general form than needed for the solution of Gelfond's problem of the sum of digits of primes.

Lemma 2.7. *Let $a > 1$ and $q = -a \pm i$. Furthermore, let $b \in \mathbb{Z}[i], \alpha \in \mathbb{R}$, $0 \leqslant \delta \leqslant \lambda$ and $k \in \mathbb{Z}[i]$ with $k \mid q^{\lambda-\delta}$ and $q \nmid k$. Then we have*

$$\sum_{\substack{h\in R_{q^\lambda}\\ h\equiv b \bmod kq^\delta}} |F_\lambda(h,\alpha)| \ll |k|^{-2\eta_5}Q^{\eta_5(\lambda-\delta)}|F_\delta(b,\alpha)|.$$

If every prime divisor p of $q = -a \pm i$ satisfies $|p|^2 \geqslant 689$, then η_5 can be replaced by η_{689}.

Proof. The proof follows the lines of the proof of [MR10, Lemma 17]. We just give a rough outline of those parts which are essentially the same as in the real case and treat the steps which are crucial in the setting of Gaussian integers in detail. If $\delta = \lambda$, then the condition $k \mid q^{\lambda-\delta}$ implies $k = 1$ and the statement holds trivially. If $\delta < \lambda$, then we define $d_\theta = (q^\theta, kq^\delta)$ (where we choose one greatest common divisor) and $u_\theta = q^\theta/d_\theta$ whenever $\delta \leqslant \theta \leqslant \lambda$. If we set $\rho_\theta = d_\theta/d_{\theta-1}$, then one can show as in the real case that the following claim holds true: ρ_θ is a Gaussian integer satisfying $\rho_\theta \mid q$ and $(q, \rho_\theta) \neq q$. Our main goal is to show that for $\delta < \theta \leqslant \lambda$,

$$\sum_{\substack{h\in R_{q^\theta}\\ h\equiv b \bmod d_\theta}} |F_\theta(h,\alpha)| \leqslant |\rho_\theta|^{-2\eta_5}Q^{\eta_5}\sum_{\substack{h\in R_{q^{\theta-1}}\\ h\equiv b \bmod d_{\theta-1}}} |F_{\theta-1}(h,\alpha)|, \tag{2.3}$$

where η_5 can be replaced by η_{689} if every prime divisor p of $q = -a \pm i$ satisfies $|p|^2 \geqslant 689$. The statement of the lemma follows then the same way as in the proof of [MR10, Lemma 17]. In order to show the result, we start with rewriting the left-hand side of (2.3). We have

$$\begin{aligned}\sum_{\substack{h\in R_{q^\theta}\\ h\equiv b \bmod d_\theta}} |F_\theta(h,\alpha)| &= \sum_{u\in R_{u_\theta}} |F_\theta(b+ud_\theta,\alpha)|\\ &= \sum_{\substack{v\in R_{qu_{\theta-1}}\\ v\equiv 0 \bmod \rho_\theta}} |F_\theta(b+vd_{\theta-1},\alpha)|\\ &= \sum_{u\in R_{u_{\theta-1}}} \sum_{\substack{w\in R_q\\ u+wu_{\theta-1}\equiv 0 \bmod \rho_\theta}} |F_\theta(b+(u+wu_{\theta-1})d_{\theta-1},\alpha)|.\end{aligned}$$

Since $u_{\theta-1}d_{\theta-1}=q^{\theta-1}$ and $F_{\theta-1}(.,\alpha)$ is periodic of period $q^{\theta-1}$, we obtain by the product representation of F_θ (see (2.1)) that

$$\sum_{\substack{h\in R_{q^\theta}\\ h\equiv b \bmod d_\theta}} |F_\theta(h,\alpha)| = \sum_{u\in R_{u_{\theta-1}}} |F_{\theta-1}(b+ud_{\theta-1},\alpha)|\; E(\theta), \tag{2.4}$$

where $E(\theta)$ is defined by

$$E(\theta)=\frac{1}{Q}\sum_{\substack{w\in R_q\\ u+wu_{\theta-1}\equiv 0 \bmod \rho_\theta}} \varphi_Q\left(\alpha-\frac{1}{2}\operatorname{tr}\left(\frac{b+ud_{\theta-1}}{q^\theta}\right)-\frac{1}{2}\operatorname{tr}\left(\frac{w}{q}\right)\right).$$

Next we seek for an upper bound of $E(\theta)$, which is the main part of this proof. Since $d_{\theta-1}(\rho_\theta,u_{\theta-1})=(d_\theta,q^{\theta-1})=(q^\theta,kq^\delta,q^{\theta-1})=d_{\theta-1}$, we see that $(\rho_\theta,u_{\theta-1})=1$. This implies that $u_{\theta-1}$ has an inverse modulo ρ_θ (say $\tilde{u}_{\theta-1}$) and we can rewrite the condition $u+wu_{\theta-1}\equiv 0 \bmod \rho_\theta$ to $w\equiv -u\tilde{u}_{\theta-1} \bmod \rho_\theta$. Since we have to consider Gaussian integers w lying in a complete residue system modulo q, we can choose $w=-u\tilde{u}_{\theta-1}-r\rho_\theta$, $r\in R_{q/\rho_\theta}$. We obtain

$$\begin{aligned}E(\theta)&=\frac{1}{Q}\sum_{r\in R_{q/\rho_\theta}} \varphi_Q\left(\alpha-\frac{1}{2}\operatorname{tr}\left(\frac{b+ud_{\theta-1}}{q^\theta}\right)-\frac{1}{2}\operatorname{tr}\left(\frac{-u\tilde{u}_{\theta-1}-r\rho_\theta}{q}\right)\right)\\ &=\frac{1}{Q}\sum_{r\in R_{q/\rho_\theta}} \varphi_Q\left(\alpha-\frac{1}{2}\operatorname{tr}\left(\frac{b+ud_{\theta-1}}{q^\theta}+\frac{u\tilde{u}_{\theta-1}}{q}\right)+\frac{1}{2}\operatorname{tr}\left(\frac{r}{q/\rho_\theta}\right)\right).\end{aligned}$$

Set $\tilde{q}=q/\rho_\theta$ and $\tilde{d}=(\Re(\tilde{q}),\Im(\tilde{q}))$. Then $\tilde{d}$ has to be a divisor of $\tilde{q}$ and thus of q. But since $q=-a\pm i$, this is only possible if $\tilde{d}=1$. Hence, $\{0\leqslant r<|\tilde{q}|^2\}$ forms a complete residue system modulo $\tilde{q}$. We get

$$E(\theta)=\frac{1}{Q}\sum_{0\leqslant r<|\tilde{q}|^2} \varphi_Q\left(\alpha-\frac{1}{2}\operatorname{tr}\left(\frac{b+ud_{\theta-1}}{q^\theta}+\frac{u\tilde{u}_{\theta-1}}{q}\right)+\frac{r\cdot\Re(\tilde{q})}{|\tilde{q}|^2}\right).$$

Note that

$$(\Re(\tilde{q}),|\tilde{q}|^2)=(\Re(\tilde{q}),\Re(\tilde{q})^2+\Im(\tilde{q})^2)=(\Re(\tilde{q}),\Im(\tilde{q})^2),$$

and that the last expression is a divisor of

$$(\Re(\tilde{q})^2, \Im(\tilde{q})^2) = (\Re(\tilde{q}), \Im(\tilde{q}))^2 = 1.$$

Thus, we have $(\Re(\tilde{q}), |\tilde{q}|^2) = 1$ and it follows that

$$E(\theta) = \frac{1}{Q} \sum_{0 \leqslant r < |\tilde{q}|^2} \varphi_Q \left(\alpha - \frac{1}{2} \operatorname{tr} \left(\frac{b + u d_{\theta-1}}{q^\theta} + \frac{u \tilde{u}_{\theta-1}}{q} \right) + \frac{r}{|\tilde{q}|^2} \right).$$

Using point (iv) of Lemma A.6 we obtain

$$E(\theta) \leqslant |\rho_\theta|^{-2\eta_{|\tilde{q}|^2}} Q^{2\eta_{|\tilde{q}|^2}}.$$

Since $\tilde{q} = q/\rho_\theta$ is a divisor of $q = -a \pm i$ we have $|\tilde{q}|^2 = 2$ or $|\tilde{q}|^2 \geqslant 5$. In both cases point (iv) (in combination with point (iii)) of Lemma A.6 implies

$$E(\theta) \leqslant |\rho_\theta|^{-2\eta_5} Q^{2\eta_5}.$$

If every prime divisor p of $q = -a \pm i$ satisfies $|p|^2 \geqslant 689$, then $|\tilde{q}|^2 \geqslant 689$ and we can replace η_5 by η_{689}. Together with (2.4) this shows the desired result. □

The last lemma of this section deals with the L^1-norm of the special case $q = -1 \pm i$. Note, that the method of the proof of Lemma 2.7 also works in this case, but then the exponent η_5 (which is smaller than 1/2, see Remark 2.9) has to be replaced by 1/2. This does not suffice for the proof of our main theorem (see for example the end of Section 2.3.3).

Lemma 2.8. *For $q = -1 \pm i$, $b \in \mathbb{Z}[i]$, $\alpha \in \mathbb{R}$ and $0 \leqslant \delta \leqslant \lambda$ we have*

$$\sum_{\substack{h \in \mathcal{F}_\lambda \\ h \equiv b \bmod q^\delta}} |F_\lambda(h, \alpha)| \leqslant 2 \cdot 2^{\eta(\lambda - \delta)} |F_\delta(b, \alpha)|,$$

where η is defined by $\eta = \frac{1}{6}\left(1 + \frac{\log(2+\sqrt{2})}{\log 2}\right)$.

Remark 2.9. Since η_5 is also given by $5^{\eta_5} = \frac{1}{5}\sum_{r=0}^{4} \left(\sin \frac{\pi}{5}\left(\frac{1}{2} + r\right)\right)^{-1}$ (see [MR10, Lemma 14]), we have $0.4272 < \eta_5 < 0.4273$. On the other hand, we have $0.4619 < \eta < 0.462$. Hence, Lemma 2.7 and Lemma 2.8 imply that for all $q = -a \pm i$ with $a \geqslant 1$ and for any $d \mid q^\lambda$ we have

$$\sum_{\substack{h \in R_{q^\lambda} \\ h \equiv b \bmod d}} |F_\lambda(h, \alpha)| \ll \left| \frac{q^\lambda}{d} \right|^{2\eta} |F_{\nu_q(d)}(b, \alpha)|,$$

where $\nu_q(d)$ denotes the unique integer k such that $q^k \mid d$ but $q^{k+1} \nmid d$.

Proof of Lemma 2.8. In order to prove this, we develop an idea of Mauduit and Rivat ([MR10, Lemma 18]). In the case $q = -1 \pm i$ the set of digits is given by the set $\{0, 1\}$ and the Fourier transform can be written as

$$|F_0(h,\alpha)| = 1, \qquad \text{and} \qquad |F_\lambda(h,\alpha)| = \prod_{j=1}^{\lambda} \left| \cos \pi \left(\alpha - \frac{1}{2} \operatorname{tr}\left(\frac{h}{q^j}\right)\right)\right|,$$

for $\lambda \geqslant 1$. Note next, that we have

$$\frac{1}{2} \operatorname{tr}\left(\frac{1}{q}\right) = -\frac{1}{2}, \qquad \frac{1}{2} \operatorname{tr}\left(\frac{1}{q^2}\right) = 0, \qquad \text{and} \qquad \frac{1}{2} \operatorname{tr}\left(\frac{1}{q^3}\right) = \frac{1}{4}, \tag{2.5}$$

for $q = -1 \pm i$ (independent of the chosen sign). We can write

$$\begin{aligned}\sum_{\substack{h \in \mathcal{F}_{\lambda+3} \\ h \equiv b \bmod q^\delta}} |F_{\lambda+3}(h,\alpha)| &= \sum_{\substack{h \in \mathcal{F}_{\lambda+2} \\ h \equiv b \bmod q^\delta}} |F_{\lambda+3}(h,\alpha)| + \sum_{\substack{h \in \mathcal{F}_{\lambda+2} \\ h \equiv b \bmod q^\delta}} |F_{\lambda+3}(h + q^{\lambda+2},\alpha)| \\ &= \sum_{\substack{h \in \mathcal{F}_{\lambda+2} \\ h \equiv b \bmod q^\delta}} |F_{\lambda+2}(h,\alpha)| \Bigg(\left| \cos \pi \left(\alpha - \frac{1}{2} \operatorname{tr}\left(\frac{h}{q^{\lambda+3}}\right)\right)\right| \\ &\qquad + \left| \sin \pi \left(\alpha - \frac{1}{2} \operatorname{tr}\left(\frac{h}{q^{\lambda+3}}\right)\right)\right| \Bigg).\end{aligned}$$

Using the estimate $|\cos x| + |\sin x| \leqslant \sqrt{2}$, we obtain that

$$\sum_{\substack{h \in \mathcal{F}_{\lambda+3} \\ h \equiv b \bmod q^\delta}} |F_{\lambda+3}(h,\alpha)| \leqslant \sqrt{2} \sum_{\substack{h \in \mathcal{F}_{\lambda+2} \\ h \equiv b \bmod q^\delta}} |F_{\lambda+2}(h,\alpha)|. \tag{2.6}$$

This does not suffice to show the desired result and we have to iterate this recurrence relation another two times. In order to get a menagable notation, we use the abbreviations

$$x = \pi \left(\alpha - \frac{1}{2} \operatorname{tr}\left(\frac{h}{q^{\lambda+1}}\right)\right), \; y = \pi \left(\alpha - \frac{1}{2} \operatorname{tr}\left(\frac{h}{q^{\lambda+2}}\right)\right), \; z = \pi \left(\alpha - \frac{1}{2} \operatorname{tr}\left(\frac{h}{q^{\lambda+3}}\right)\right).$$

Observing (2.5), we get

$$\begin{aligned}&\sum_{\substack{h \in \mathcal{F}_{\lambda+3} \\ h \equiv b \bmod q^\delta}} |F_{\lambda+3}(h,\alpha)| \\ &\quad = \sum_{\substack{h \in \mathcal{F}_{\lambda} \\ h \equiv b \bmod q^\delta}} |F_{\lambda}(h,\alpha)| \Big(|\cos x| (|\cos y| + |\sin y|)(|\cos z| + |\sin z|) \\ &\qquad\qquad + |\sin x| (|\cos y| + |\sin y|)(|\cos(z - \pi/4)| + |\sin(z - \pi/4)|) \Big).\end{aligned}$$

Using $(|\cos x| + |\sin x|)^2 = 1 + |\sin 2x|$, $|\cos x| + |\sin x| \leqslant \sqrt{2}$ as well as

$$(|\cos\theta|a + |\sin\theta|b)^2 \leqslant a^2 + b^2,$$

we obtain

$$\sum_{\substack{h\in\mathcal{F}_{\lambda+3}\\ h\equiv b \bmod q^\delta}} |F_{\lambda+3}(h,\alpha)| \leqslant \sqrt{2}\sqrt{2+\sqrt{2}} \sum_{\substack{h\in\mathcal{F}_{\lambda}\\ h\equiv b \bmod q^\delta}} |F_{\lambda}(h,\alpha)|.$$

Applying this inequality $\lfloor\frac{\lambda-\delta}{3}\rfloor$-times (and estimate (2.6) one or two times, if needed) yields the desired result. □

2.3 Proof of Theorem 2.1

In this section we prove Theorem 2.1, that is, we give an upper bound of the exponential sum

$$\sum_{\substack{|n|^2\leqslant N\\ \gamma_1\leqslant \mathrm{Arg}(n)<\gamma_2}} \Lambda_i(n)\,\mathrm{e}(\alpha s_q(n)). \tag{2.7}$$

2.3.1 Vaughan's identity

In order to handle the sum (2.7), we decompose the von Mangoldt function into a sum of other functions. In particular, we use a variant of Vaughan's identity in the Gaussian integers (similar to [DRS08, Lemma 3.1]).

Lemma 2.10. *Let $Q \geqslant 2$ be an integer, $0 < \beta_1 < \frac{1}{3}$, $\frac{1}{2} < \beta_2 < 1$ real numbers and g an arithmetic function. Suppose that uniformly for all real numbers $M \leqslant x$ and all complex numbers a_m, b_n with $|a_m|, |b_n| \leqslant 1$, we have*

$$\sum_{\frac{M}{Q}<|m|^2\leqslant M} \max_{\frac{x}{Q|m|^2}<t\leqslant\frac{x}{|m|^2}} \left| \sum_{\substack{\frac{x}{Q|m|^2}<|n|^2\leqslant t\\ \gamma_1\leqslant\mathrm{Arg}(mn)<\gamma_2}} g(mn)\right| \leqslant U \quad \textit{for } M\leqslant x^{\beta_1} \quad \textit{(type I)}, \tag{2.8}$$

$$\left| \sum_{\frac{M}{Q}<|m|^2\leqslant M} \sum_{\substack{\frac{x}{Q|m|^2}<|n|^2\leqslant\frac{x}{|m|^2}\\ \gamma_1\leqslant\mathrm{Arg}(mn)<\gamma_2}} a_m b_n g(mn)\right| \leqslant U \quad \textit{for } x^{\beta_1}\leqslant M\leqslant x^{\beta_2} \quad \textit{(type II)}. \tag{2.9}$$

Then

$$\sum_{\substack{\frac{x}{Q}<|n|^2\leqslant x\\ \gamma_1\leqslant\mathrm{Arg}(mn)<\gamma_2}} \Lambda_i(n) g(n) \ll_q U(\log x)^2.$$

Remark 2.11. Drmota et al. used Dirichlet series and Dedekind's zeta function of $\mathbb{Q}(i)$ for proving the decomposition of $\sum_n \Lambda_i(n)g(n)$ into three different sums (cf. (2.10)). Without further reasonings this seems not to be justified since two Dirichlet series

$$\sum_{n\in\mathbb{Z}[i]\setminus\{0\}} \frac{a_n}{|n|^{2s}} \qquad \text{and} \qquad \sum_{n\in\mathbb{Z}[i]\setminus\{0\}} \frac{b_n}{|n|^{2s}}$$

can be equal even if $a_n \neq b_n$ for some $n \in \mathbb{Z}[i]$. (Take for example the Dirichlet series defined by $a_1 = 1$ and $a_n = 0$ for $n \neq 1$ and $b_i = 1$ and $b_n = 0$ for $n \neq i$. Both series are given by the constant function 1.) In particular, one cannot deduce the coefficients a_n from the corresponding Dirichlet series $\sum_n a_n/|n|^{2s}$. Thus, we give a proof of the performed decomposition without using Dirichlet series in the Gaussian integers (compare with [IK04, Chapter 13.4]).

Proof of Lemma 2.10. Let $1 \leqslant u \leqslant x/Q < |n|^2 \leqslant x$. Then we have

$$\frac{1}{2} \cdot \sum_{\substack{\frac{x}{Q}<|n|^2\leqslant x\\ \gamma_1\leqslant \operatorname{Arg}(n)<\gamma_2}} \Lambda_i(n)g(n) = S_1 - S_2 + S_3, \tag{2.10}$$

where

$$S_1 = \frac{1}{8} \cdot \sum_{|m|^2\leqslant u} \sum_{\substack{\frac{x}{Q|m|^2}<|n|^2\leqslant \frac{x}{|m|^2}\\ \gamma_1\leqslant \operatorname{Arg}(mn)<\gamma_2}} \mu_i(m)\log(|n|)g(mn),$$

$$S_2 = \frac{1}{32} \cdot \sum_{|m_1|^2,|m_2|^2\leqslant u} \sum_{\substack{\frac{x}{Q|m_1m_2|^2}<|n|^2\leqslant \frac{x}{|m_1m_2|^2}\\ \gamma_1\leqslant \operatorname{Arg}(m_1m_2n)<\gamma_2}} \mu_i(m_1)\Lambda_i(m_2)g(m_1m_2n),$$

$$S_3 = \frac{1}{32} \cdot \sum_{|m|^2,|n_1|^2> u} \sum_{\substack{\frac{x}{Q|mn_1|^2}<|n_2|^2\leqslant \frac{x}{|mn_1|^2}\\ \gamma_1\leqslant \operatorname{Arg}(mn_1n_2)<\gamma_2}} \mu_i(m)\Lambda_i(n_1)g(mn_1n_2),$$

and μ_i denotes the complex Möbius function

$$\mu_i(n) = \begin{cases} 1, & \text{if } n \in \{\pm 1, \pm i\},\\ (-1)^\nu, & \text{if } n \text{ is a product of } \nu \text{ distinct primes (times some unit)},\\ 0, & \text{otherwise.} \end{cases}$$

In order to show this, we need the following formulas (which are slightly different from the appropriate formulas of the common von Mangoldt and Möbius function). We have

$$\sum_{m|n} \Lambda_i(m) = 4\log|n| \tag{2.11}$$

and

$$\sum_{m|n} \mu_i(m) = \begin{cases} 4, & \text{if } n \in \{\pm 1, \pm i\}, \\ 0, & \text{otherwise.} \end{cases} \tag{2.12}$$

Let us fix some Gaussian integer n. We start with the relation

$$16\Lambda_i(n) = \sum_{\substack{b,c \\ bc|n}} \Lambda_i(b)\mu_i(c).$$

It is a direct consequence of (2.12) and the fact that $\Lambda_i(n) = \Lambda_i(-n) = \Lambda_i(in) = \Lambda_i(-in)$. Next, we split the sum up according to the size of b and c to obtain

$$\begin{aligned} 16\Lambda_i(n) = & \sum_{\substack{|b|^2 \leqslant u,\, |c|^2 \leqslant u \\ bc|n}} \Lambda_i(b)\mu_i(c) + \sum_{\substack{|b|^2 \leqslant u,\, |c|^2 > u \\ bc|n}} \Lambda_i(b)\mu_i(c) \\ & + \sum_{\substack{|b|^2 > u,\, |c|^2 \leqslant u \\ bc|n}} \Lambda_i(b)\mu_i(c) + \sum_{\substack{|b|^2 > u,\, |c|^2 > u \\ bc|n}} \Lambda({}_ib)\mu_i(c). \end{aligned}$$

Moreover, we have

$$\sum_{\substack{|b|^2 \leqslant u,\, |c|^2 \leqslant u \\ bc|n}} \Lambda_i(b)\mu_i(c) + \sum_{\substack{|b|^2 \leqslant u,\, |c|^2 > u \\ bc|n}} \Lambda_i(b)\mu_i(c) = \sum_{\substack{|b|^2 \leqslant u \\ b|n}} \Lambda_i(b) \sum_{c|\frac{n}{b}} \mu_i(c),$$

which is equal to zero, since by assumption $|n|^2 > u$ (see formula (2.12)). Furthermore, we obtain (using (2.11))

$$\begin{aligned} \sum_{\substack{|b|^2 \leqslant u,\, |c|^2 \leqslant u \\ bc|n}} \Lambda_i(b)\mu_i(c) + \sum_{\substack{|b|^2 > u,\, |c|^2 \leqslant u \\ bc|n}} \Lambda_i(b)\mu_i(c) &= \sum_{\substack{|c|^2 \leqslant u \\ c|n}} \mu_i(c) \sum_{b|\frac{n}{c}} \Lambda_i(b) \\ &= 4 \sum_{\substack{|c|^2 \leqslant u \\ c|n}} \mu_i(c) \log \left| \frac{n}{c} \right|. \end{aligned}$$

Hence, taking these facts into account, we get ($|n|^2 > u$)

$$16\Lambda_i(n) = 4 \sum_{\substack{|c|^2 \leqslant u \\ c|n}} \mu_i(c) \log \left| \frac{n}{c} \right| - \sum_{\substack{|b|^2 \leqslant u,\, |c|^2 \leqslant u \\ bc|n}} \Lambda_i(b)\mu_i(c) + \sum_{\substack{|b|^2 > u,\, |c|^2 > u \\ bc|n}} \Lambda_i(b)\mu_i(c).$$

Weighting by $g(n)$, dividing by 32 and summing over all n with $x/Q < |n|^2 \leqslant x$ and $\gamma_1 \leqslant \operatorname{Arg}(n) < \gamma_2$ we finally get (2.10).

The rest of the proof is similar to the rest of the proof of [DRS08, Lemma 3.1]. We choose $u = x^{\beta_1}$. In order to bound S_1, we use partial summation to obtain

$$S_1 \leqslant \frac{1}{8} \cdot \sum_{|m|^2 \leqslant u} \left| \sum_{\substack{\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2} \\ \gamma_1 \leqslant \operatorname{Arg}(mn) < \gamma_2}} \log(|n|) g(mn) \right|$$

$$\ll (\log x) \sum_{|m|^2 \leqslant u} \max_{\frac{x}{Q|m|^2} < t \leqslant \frac{x}{|m|^2}} \left| \sum_{\substack{\frac{x}{Q|m|^2} < |n|^2 \leqslant t \\ \gamma_1 \leqslant \operatorname{Arg}(mn) < \gamma_2}} g(mn) \right| .$$

Splitting the summation over m gives

$$S_1 \ll (\log x)^2 \max_{M \leqslant u} \sum_{\frac{M}{Q} < |m|^2 \leqslant M} \max_{\frac{x}{Q|m|^2} < t \leqslant \frac{x}{|m|^2}} \left| \sum_{\substack{\frac{x}{Q|m|^2} < |n|^2 \leqslant t \\ \gamma_1 \leqslant \operatorname{Arg}(mn) < \gamma_2}} g(mn) \right| .$$

The type I estimate (2.8) implies that $S_1 \ll_q U(\log x)^2$. For the estimation of S_2 we use (see (2.11))

$$\left| \sum_{\substack{|m_1|^2, |m_2|^2 \leqslant u \\ m = m_1 m_2}} \mu_i(m_1) \Lambda_i(m_2) \right| \leqslant \sum_{d|m} \Lambda_i(d) = 4 \log |m|,$$

to obtain

$$S_2 \ll (\log x)^2 \max_{M \leqslant u^2} \sum_{\frac{M}{Q} < |m|^2 \leqslant M} \left| \sum_{\substack{\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2} \\ \gamma_1 \leqslant \operatorname{Arg}(mn) < \gamma_2}} g(mn) \right| .$$

Let M_0 be a value of M for which the maximum is attained. If $M_0 \leqslant u$ $(= x^{\beta_1})$ or $u < M_0 \leqslant x^{\frac{1}{2}}$ we can employ (2.8) in the first case or (2.9) in the second case to derive $S_2 \ll U(\log x)^2$. In the case that $x^{\frac{1}{2}} < M_0 \leqslant u^2$ we can choose complex numbers a_m such that

$$\sum_{\frac{M_0}{Q} < |m|^2 \leqslant M_0} \left| \sum_{\substack{\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2} \\ \gamma_1 \leqslant \operatorname{Arg}(mn) < \gamma_2}} g(mn) \right| = \sum_{\frac{M_0}{Q} < |m|^2 \leqslant M_0} \sum_{\substack{\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2} \\ \gamma_1 \leqslant \operatorname{Arg}(mn) < \gamma_2}} a_m g(mn).$$

Setting $a_m = 0$ if $|m|^2 > M_0$ or $|m|^2 \leqslant M_0/Q$, we are able to change the order of summation and get

$$\begin{aligned}
&\sum_{\frac{M_0}{Q}<|m|^2\leqslant M_0} \left| \sum_{\substack{\frac{x}{Q|m|^2}<|n|^2\leqslant\frac{x}{|m|^2}\\ \gamma_1\leqslant\mathrm{Arg}(mn)<\gamma_2}} g(mn) \right| \\
&= \sum_{\frac{x}{M_0 Q}<|n|^2\leqslant\frac{x}{M_0}} \sum_{\substack{\frac{x}{Q|n|^2}<|m|^2\leqslant\frac{x}{|n|^2}\\ \gamma_1\leqslant\mathrm{Arg}(mn)<\gamma_2}} a_m g(mn) + \sum_{\frac{x}{M_0}<|n|^2\leqslant\frac{xQ}{M_0}} \sum_{\substack{\frac{x}{Q|n|^2}<|m|^2\leqslant\frac{x}{|n|^2}\\ \gamma_1\leqslant\mathrm{Arg}(mn)<\gamma_2}} a_m g(mn).
\end{aligned}$$

If we define $M_1 = \frac{x}{M_0}$ and $M_2 = \frac{xQ}{M_0}$, we have for sufficiently large x that $x^{\beta_1} \leqslant M_1 \leqslant x^{\beta_2}$ and $x^{\beta_1} \leqslant M_2 \leqslant x^{\beta_2}$. Thus we can employ the type II estimation (2.9) to the first sum with $M = M_1$ and to the second sum with $M = M_2$ and we obtain $S_2 \ll_q U(\log x)^2$ in this case, too. To bound S_3 we write

$$S_3 = \frac{2\log x}{32} \sum_{u<|m|^2\leqslant\frac{x}{u}} \sum_{\substack{\frac{x}{Q|m|^2}<|n|^2\leqslant\frac{x}{|m|^2}\\ \gamma_1\leqslant\mathrm{Arg}(mn)<\gamma_2}} a_m b_n g(mn),$$

where

$$a_m = \mu_i(m) \qquad \text{and} \qquad b_n = \frac{1}{2\log x} \sum_{\substack{u<|n_1|^2,\,|n_2|^2<|n|^2\\ n=n_1 n_2}} \Lambda_i(n_1).$$

These numbers satisfy $|a_m| \leqslant 1$ and $0 \leqslant b_n \leqslant \frac{1}{2\log x}\sum_{d|n}\Lambda_i(d) = \frac{4\log|n|}{2\log x} \leqslant 1$. Splitting up the summation according to the powers of Q, we obtain

$$S_3 \ll (\log x)^2 \max_{u\leqslant M\leqslant\frac{x}{u}} \left| \sum_{\frac{M}{Q}<|m|^2\leqslant M} \sum_{\substack{\frac{x}{Q|m|^2}<|n|^2\leqslant\frac{x}{|m|^2}\\ \gamma_1\leqslant\mathrm{Arg}(mn)<\gamma_2}} a_m b_n g(mn) \right|.$$

We carry out a similar procedure as before. If the maximum M_0 is attained between x^{β_1} and $x^{1/2}$, we can directly use the type II sum estimate. Otherwise we have to change the order of summation and use then the type II sum estimate. We finally obtain $S_3 \ll_q U(\log x)^2$, which implies the desired result. □

2.3.2 Sums of type I

The type I sums for “discs” (that is, for $\gamma_1 = 0$ and $\gamma_2 = 1$) are treated by Drmota et al. in [DRS08, Chapter 8] for $a \geqslant 2$. Note, that the assumption $a \neq 1$ is not needed in their reasoning so that they actually proved the type I sum estimate for all $a \geqslant 1$. In what follows we show that this estimate is also valid for general angles γ_1 and γ_2. Since the proof is very similar to the proof of Drmota et al., we give just an outline and refer at appropriate places to their work.

Proposition 2.12. *Let $a \in \mathbb{Z}^+$, $q = -a \pm i$, $0 \leqslant \gamma_1 < \gamma_2 \leqslant 1$ and α a real number such that $(a^2 + 2a + 2)\alpha \in \mathbb{R} \setminus \mathbb{Z}$. Then there exists a positive constant $\sigma'_q(\alpha)$, such that for all M with $1 \leqslant M \leqslant x^{1/10}$ we have*

$$\sum_{\frac{M}{Q} < |m|^2 \leqslant M} \max_{\frac{x}{Q|m|^2} < t \leqslant \frac{x}{|m|^2}} \left| \sum_{\substack{\frac{x}{Q|m|^2} < |n|^2 \leqslant t \\ \gamma_1 \leqslant \operatorname{Arg}(mn) < \gamma_2}} \mathrm{e}\left(\alpha s_q(mn)\right) \right| \ll_{q,\alpha} x^{1-\sigma'_q(\alpha)}.$$

Proof. In order to prove this proposition it suffices to show that

$$\sum_{\frac{M}{Q} < |m|^2 \leqslant M} \left| \sum_{\substack{|n|^2 \leqslant t \\ \gamma_1 \leqslant \operatorname{Arg}(mn) < \gamma_2}} \mathrm{e}\left(\alpha s_q(mn)\right) \right| \ll_{q,\alpha} x^{1-\sigma'_q(\alpha)} \tag{2.13}$$

for all $\frac{x}{QM} \leqslant t \leqslant \frac{xQ}{M}$. If R_m is a complete residue system modulo m (whose elements k satisfy $|k| \leqslant (\sqrt{2}/2)|m|$), then we have

$$\left| \sum_{\substack{|n|^2 \leqslant t \\ \gamma_1 \leqslant \operatorname{Arg}(mn) < \gamma_2}} \mathrm{e}\left(\alpha s_q(mn)\right) \right| = \left| \frac{1}{|m|^2} \sum_{k \in R_m} \sum_{\substack{|\ell|^2 \leqslant |m|^2 t \\ \gamma_1 \leqslant \operatorname{Arg}(\ell) < \gamma_2}} \mathrm{e}\left(\alpha s_q(\ell) + \frac{1}{2} \operatorname{tr} \frac{k\ell}{m}\right) \right|.$$

Note, that $x/Q^2 < |m|^2 t \leqslant xQ$ and that the sum-of-digits function is completely q-additive (that is, we have $s_q(kq^\lambda + \ell) = s_q(k) + s_q(\ell)$ for all $\lambda \in \mathbb{Z}^+$, $k \in \mathbb{Z}[i]$ and $\ell \in \mathcal{F}_\lambda$). We tessellate the domain $\{\ell \in \mathbb{Z}[i] : |\ell|^2 \leqslant |m|^2 t, \gamma_1 \leqslant \operatorname{Arg}(\ell) < \gamma_2\}$ by translates of $\mathcal{F}_\lambda$, where $\lambda = \left\lfloor \frac{3 \log x}{10 \log Q} \right\rfloor$. Here is a difference to the treatment in [DRS08] since they consider the whole disc $\{\ell \in \mathbb{Z}[i] : |\ell|^2 \leqslant |m|^2 t\}$ instead of a circular sector. However, the reasoning is quite similar and we obtain

$$\left| \sum_{\substack{|\ell|^2 \leqslant |m|^2 t \\ \gamma_1 \leqslant \operatorname{Arg}(\ell) < \gamma_2}} \mathrm{e}\left(\alpha s_q(\ell) + \frac{1}{2} \operatorname{tr} \frac{k\ell}{m}\right) \right| \ll \frac{x}{Q^\lambda} \left| \sum_{\ell \in \mathcal{F}_\lambda} \mathrm{e}\left(\alpha s_q(\ell) + \frac{1}{2} \operatorname{tr} \frac{k\ell}{m}\right) \right| + O(x^{1/2}|q|^\lambda). \tag{2.14}$$

Using a two-dimensional large sieve based on the Sobolev-Gallagher inequality, it is shown in [DRS08, Lemma 8.1 and Proposition 8.2] that

$$\sum_{\frac{M}{Q} < |m|^2 \leqslant M} \sum_{k \in R_m} \left| \sum_{\ell \in \mathcal{F}_\lambda} \mathrm{e}\left(\alpha s_q(\ell) + \frac{1}{2} \operatorname{tr} \frac{k\ell}{m}\right) \right| \ll_q M^2 Q^{\lambda/2} + Q^{\gamma'_Q(\alpha)\lambda} M^{3-2\gamma'_Q(\alpha)},$$

where $0 \leqslant \gamma'_Q(\alpha) < 1$. One can now prove (exactly the same way as at the end of [DRS08, Section 8]) that this result together with (2.14) implies the existence of a constant $\sigma'_q(\alpha)$ with $\sigma'_q(\alpha) < 1/20 \min(1, 2 - 2\gamma'_Q(\alpha))$ such that (2.13) holds true for all $\frac{x}{QM} \leqslant t \leqslant \frac{xQ}{M}$. □

2.3.3 Sums of type II

Proposition 2.13. *Let $a \in \mathbb{Z}^+$, $q = -a \pm i$, $0 \leqslant \gamma_1 < \gamma_2 \leqslant 1$ and α a real number such that $(a^2+2a+2)\alpha \in \mathbb{R} \setminus \mathbb{Z}$. Then there exists a positive constant $\sigma''_q(\alpha)$ and real numbers $0 < \beta_1 < 1/10$ and $1/2 < \beta_2 < 2/3$ such that the following holds true: For any complex numbers a_m, b_n with $|a_m|$, $|b_n| \leqslant 1$ and for all M with $x^{\beta_1} \leqslant M \leqslant x^{\beta_2}$ we have*

$$\left| \sum_{\frac{M}{Q} < |m|^2 \leqslant M} \sum_{\substack{\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2} \\ \gamma_1 \leqslant \operatorname{Arg}(mn) < \gamma_2}} a_m b_n g(mn) \right| \ll_{q,\alpha} (\log x)^2 x^{1-\sigma''_q(\alpha)}. \tag{2.15}$$

In the next lemma, we "remove" the dependency of m in the domain of summation over n. In contrast to [DRS08], we also have to deal with the fact that mn lies in an circular sector.

Lemma 2.14. *Suppose that there exist real numbers $\beta_1 \in \left(0, \frac{1}{3}\right)$, $\beta_2 \in \left(\frac{1}{2}, 2/3\right)$ and $0 < \hat{\sigma}''_q(\alpha) < 1$, only depending on α and q, such that for any complex numbers b_n with $|b_n| \leqslant 1$ and for all M with $x^{\beta_1} \leqslant M \leqslant x^{\beta_2}$ we have*

$$\sum_{\frac{M}{Q} < |m|^2 \leqslant M} \left| \sum_{\frac{x}{QM} < |n|^2 \leqslant \frac{xQ}{M}} b_n g(mn) \right| \ll_{q,\alpha} x^{1-\hat{\sigma}''_q(\alpha)}. \tag{2.16}$$

Then the type II sum estimate (2.15) *holds true with* $\sigma''_q(\alpha) = \min(\hat{\sigma}''_q(\alpha), 1/6)$.

First Proof of Lemma 2.14. In what follows, we show that Lemma 2.14 holds true with $\sigma''_q(\alpha) = \min(\hat{\sigma}''_q(\alpha), (1-\beta_2)/2) \geqslant \min(\hat{\sigma}''_q(\alpha), 1/6)$. We start with removing the condition that mn lies in a circular sector. Since $\operatorname{Arg}(mn) = \operatorname{Arg}(m) + \operatorname{Arg}(n) \bmod 1$ and $0 \leqslant \operatorname{Arg}(m), \operatorname{Arg}(n) < 1$, we have $\gamma_1 \leqslant \operatorname{Arg}(mn) < \gamma_2$ if and only if $\gamma_1 \leqslant \operatorname{Arg}(m) + \operatorname{Arg}(n) < \gamma_2$ or $\gamma_1 \leqslant \operatorname{Arg}(m) + \operatorname{Arg}(n) - 1 < \gamma_2$ (where both conditions cannot be satisfied simultaneously). This is equivalent to the fact that $\gamma_1 \leqslant \operatorname{Arg}(mn) < \gamma_2$ if and only if $e^{-\gamma_1} \geqslant e^{-\operatorname{Arg}(m) - \operatorname{Arg}(n)} > e^{-\gamma_2}$ or $e^{-\gamma_1} \geqslant e^{-\operatorname{Arg}(m) - \operatorname{Arg}(n)+1} > e^{-\gamma_2}$ (again, both conditions cannot be satisfied simultaneously). We have

$$\begin{aligned} \sum_{\frac{M}{Q} < |m|^2 \leqslant M} \left| \sum_{\substack{\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2} \\ \gamma_1 \leqslant \operatorname{Arg}(mn) < \gamma_2}} b_n g(mn) \right| &\leqslant \sum_{\frac{M}{Q} < |m|^2 \leqslant M} \left| \sum_{\substack{\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2} \\ e^{-\gamma_1} \geqslant e^{-\operatorname{Arg}(m) - \operatorname{Arg}(n)} > e^{-\gamma_2}}} b_n g(mn) \right| \\ &+ \sum_{\frac{M}{Q} < |m|^2 \leqslant M} \left| \sum_{\substack{\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2} \\ e^{-\gamma_1} \geqslant e^{-\operatorname{Arg}(m) - \operatorname{Arg}(n)+1} > e^{-\gamma_2}}} b_n g(mn) \right|. \end{aligned} \tag{2.17}$$

In the following, we show that the first term on the right-hand side of (2.17) is bounded above by $(\log x)^2 x^{1-\sigma_q''(\alpha)}$. In almost the same way one can show that the second term is also bounded by the same value which then implies the desired result. Set $f(t) := \mathbf{1}_{(e^{-\gamma_2}, e^{-\gamma_1}]}(t)$, where $\mathbf{1}_A$ denotes the characteristic function of the set A. We can write

$$\left| \sum_{\substack{\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2} \\ e^{-\gamma_1} \geqslant e^{-\operatorname{Arg}(m) - \operatorname{Arg}(n)} > e^{-\gamma_2}}} b_n g(mn) \right| = \left| \sum_{\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2}} b_n g(mn) f\left(e^{-\operatorname{Arg}(m) - \operatorname{Arg}(n)} \right) \right|.$$

It follows from Perron's formula (see for example [Ten08, Section 2.1] and [FI89, Lemma 6]), that

$$\left| f(t) - \frac{1}{2\pi i} \int_{-L}^{L} t^{-is} \frac{e^{-\gamma_1 is} - e^{-\gamma_2 is}}{s} \mathrm{d}s \right| \ll L^{-1/2} \tag{2.18}$$

for all $t \geqslant 0$ with $\min(|t - e^{-\gamma_1}|, |t - e^{-\gamma_2}|) \geqslant L^{-1/2}$. If $e^{-\operatorname{Arg}(m) - \operatorname{Arg}(n)}$ is not too close to $e^{-\gamma_1}$ and $e^{-\gamma_2}$, we will apply this approximation with $L = x$. For all other Gaussian integers we use a trivial estimate to bound the considered sum. Let $I(m, x)$ be the set of Gaussian integers n with $\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2}$ and

$$\min\left(\left| e^{-\operatorname{Arg}(m) - \operatorname{Arg}(n)} - e^{-\gamma_1} \right|, \left| e^{-\operatorname{Arg}(m) - \operatorname{Arg}(n)} - e^{-\gamma_2} \right| \right) < x^{-1/2}.$$

Elementary transformations show that if $n \in I(m, x)$, then

$$\gamma_1 - \operatorname{Arg}(m) - e^{\gamma_1} x^{-1/2} < \operatorname{Arg}(n) < \gamma_1 - \operatorname{Arg}(m) + 3e^{\gamma_1} x^{-1/2}$$

or

$$\gamma_2 - \operatorname{Arg}(m) - e^{\gamma_2} x^{-1/2} < \operatorname{Arg}(n) < \gamma_2 - \operatorname{Arg}(m) + 3e^{\gamma_2} x^{-1/2}.$$

This implies that for $\frac{M}{Q} < |m|^2 \leqslant M$ we have $\# I(m, x) \ll x^{1/2}/|m| \ll_q x^{1/2}/M^{1/2}$. Indeed, this follows from the fact that $I(m, x)$ is contained in a narrow circular sector and that $|n| \leqslant x^{1/2}/|m|$ for all $n \in I(m, x)$. We obtain

$$\left| \sum_{\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2}} b_n g(mn) f\left(e^{-\operatorname{Arg}(m) - \operatorname{Arg}(n)} \right) \right| \\ \ll_q \left| \sum_{\substack{\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2} \\ n \notin I(m,x)}} b_n g(mn) f\left(e^{-\operatorname{Arg}(m) - \operatorname{Arg}(n)} \right) \right| + \frac{x^{1/2}}{M^{1/2}}. \tag{2.19}$$

Using relation (2.18) with $L = x$, the sum on the right-hand side of (2.19) can be bounded by

$$\left| \frac{1}{2\pi i} \int_{-x}^{x} e^{is \operatorname{Arg}(m)} \left(\sum_{\substack{\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2} \\ n \notin I(m,x)}} b_n g(mn) e^{is \operatorname{Arg}(n)} \right) \frac{e^{-\gamma_1 is} - e^{-\gamma_2 is}}{s} \mathrm{d}s \right| + \frac{xQ}{M} x^{-1/2}.$$

This yields

$$\left| \sum_{\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2}} b_n g(mn) f\left(e^{-\operatorname{Arg}(m) - \operatorname{Arg}(n)} \right) \right|$$
$$\ll_q \frac{1}{2\pi} \int_{-x}^{x} \left| \sum_{\substack{\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2} \\ n \notin I(m,x)}} b_n g(mn) e^{is \operatorname{Arg}(n)} \right| \frac{\left| e^{-\gamma_1 is} - e^{-\gamma_2 is} \right|}{|s|} \mathrm{d}s + \frac{x^{1/2}}{M^{1/2}}.$$

Including the Gaussian integers $n \in I(m,x)$ in the last formula yields an additional error term

$$\frac{1}{2\pi} \int_{-x}^{x} \left| \sum_{n \in I(m,x)} b_n g(mn) e^{is \operatorname{Arg}(n)} \right| \frac{\left| e^{-\gamma_1 is} - e^{-\gamma_2 is} \right|}{|s|} \mathrm{d}s$$
$$\ll_q \int_{-x}^{x} \frac{x^{1/2}}{M^{1/2}} \frac{\left| e^{-\gamma_1 is} - e^{-\gamma_2 is} \right|}{|s|} \mathrm{d}s.$$

Since $(e^{-\gamma_1 is} - e^{-\gamma_2 is})/s$ is bounded in a neighborhood of zero, we see that

$$\int_{-x}^{x} \frac{\left| e^{-\gamma_1 is} - e^{-\gamma_2 is} \right|}{|s|} \mathrm{d}s \ll \log x. \tag{2.20}$$

Hence, this error term is $\ll_q (\log x)\, x^{1/2}/M^{1/2}$. We obtain

$$\sum_{\frac{M}{Q} < |m|^2 \leqslant M} \left| \sum_{\substack{\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2} \\ e^{-\gamma_1} \geqslant e^{-\operatorname{Arg}(m) - \operatorname{Arg}(n)} > e^{-\gamma_2}}} b_n g(mn) \right|$$
$$\ll_q \int_{-x}^{x} \sum_{\frac{M}{Q} < |m|^2 \leqslant M} \left| \sum_{\frac{x}{Q|m|^2} < |n|^2 \leqslant \frac{x}{|m|^2}} b_n g(mn) e^{is \operatorname{Arg}(n)} \right| \frac{\left| e^{-\gamma_1 is} - e^{-\gamma_2 is} \right|}{|s|} \mathrm{d}s + E(x, M), \tag{2.21}$$

where $E(x,M)=(\log x)x^{1/2}M^{1/2}\leqslant(\log x)x^{1-\sigma''_q(\alpha)}$. Hence, it remains to bound the considered integral. In order to be able to apply (2.16), we use the same method as in [DRS08, Proposition 3.2] to eliminate the dependence of m in the domain of summation over n. We have

$$\begin{aligned}&\sum_{\frac{x}{Q|m|^2}<|n|^2\leqslant\frac{x}{|m|^2}} b_n g(mn) e^{is\operatorname{Arg}(n)}\\ &\quad=\int_{-1/2}^{1/2}\left(\sum_{\substack{\frac{x}{QM}<|n|^2\leqslant\frac{xQ}{M}\\ n\in\mathbb{Z}[i]}} b_n g(mn)e^{is\operatorname{Arg}(n)}\,\mathrm{e}(|n|^2\xi)\right)\cdot\left(\sum_{\substack{\frac{x}{Q|m|^2}<n'\leqslant\frac{x}{|m|^2}\\ n'\in\mathbb{Z}}}\mathrm{e}(-n'\xi)\right)\mathrm{d}\xi.\end{aligned}$$

Estimating the geometric series and taking the absolute value yields an upper bound of the form

$$\int_{-1/2}^{1/2}\min(x,|\sin\pi\xi|^{-1})\left|\sum_{\frac{x}{QM}<|n|^2\leqslant\frac{xQ}{M}} b_n\,\mathrm{e}(|n|^2\xi)g(mn)e^{is\operatorname{Arg}(n)}\right|\mathrm{d}\xi.$$

Since $\int_{-1/2}^{1/2}\min(x,|\sin\pi\xi|^{-1})\mathrm{d}\xi\ll\log x$ (see for example [MR10, Lemma 2]) and since estimate (2.16) is uniform in b_n (take $b_n\,\mathrm{e}(|n|^2\xi)e^{is\operatorname{Arg}(n)}$ instead of b_n), we obtain

$$\sum_{\frac{M}{Q}<|m|^2\leqslant M}\left|\sum_{\frac{x}{Q|m|^2}<|n|^2\leqslant\frac{x}{|m|^2}} b_n g(mn)e^{is\operatorname{Arg}(n)}\right|\ll_q(\log x)x^{1-\hat\sigma''_q(\alpha)}\ll_q(\log x)x^{1-\sigma''_q(\alpha)}.$$

This together with (2.20) and (2.21) finally implies the desired result. □

Second Proof of Lemma 2.14. As in the first proof, we start with removing the condition that mn lies in a circular sector. This time, we define the function f in a slightly different way, namely $f(t):=\mathbf{1}_{[\gamma_1,\gamma_2)}(\{t\})$, where $\mathbf{1}_A$ again denotes the characteristic function of the set A. The function f is periodic and we approximate it by trigonometric polynomials. Let $H\geqslant1$ be an integer chosen later on. Then there exist coefficients $a_H(h)$ with $|a_H(h)|\leqslant2$, such that

$$f_H^*(t)=(\gamma_2-\gamma_1)+\frac{1}{2\pi i}\sum_{1\leqslant|h|\leqslant H}\frac{a_H(h)}{h}\,\mathrm{e}(ht)$$

and

$$\kappa_H(t)=\sum_{|h|\leqslant H}\left(1-\frac{|h|}{H+1}\right)\mathrm{e}(ht)$$

verify

$$|f(t) - f_H^*(t)| \leqslant \frac{1}{2H+2}\Big(\kappa_H\,(t-\gamma_1) + \kappa_H\,(t-\gamma_2)\Big).$$

This follows from Vaaler's approximation method using the Beurling-Selberg function (see CorollaryA.22). We get

$$\begin{aligned}\Bigg|\sum_{\substack{\frac{x}{Q|m|^2}<|n|^2\leqslant\frac{x}{|m|^2}\\ \gamma_1\leqslant\operatorname{Arg}(mn)<\gamma_2}} b_n g(mn)\Bigg| &= \Bigg|\sum_{\frac{x}{Q|m|^2}<|n|^2\leqslant\frac{x}{|m|^2}} b_n g(mn) f(\operatorname{Arg}(mn))\Bigg|\\ \leqslant \Bigg|&\sum_{\frac{x}{Q|m|^2}<|n|^2\leqslant\frac{x}{|m|^2}} b_n g(mn)\left((\gamma_2-\gamma_1) + \frac{1}{2i\pi}\sum_{1\leqslant|h|\leqslant H}\frac{a_H(h)}{h}\,\mathrm{e}(h\operatorname{Arg}(mn))\right)\Bigg|\\ + &\sum_{\frac{x}{Q|m|^2}<|n|^2\leqslant\frac{x}{|m|^2}} \frac{1}{2H+2}\Big(\kappa_H(\operatorname{Arg}(mn)-\gamma_1) + \kappa_H(\operatorname{Arg}(mn)-\gamma_2)\Big).\end{aligned}$$

Note, that $\operatorname{Arg}(mn) = \operatorname{Arg}(m) + \operatorname{Arg}(n) \bmod 1$. Thus, if we additionally sum over m, we obtain

$$\sum_{\frac{M}{Q}<|m|^2\leqslant M}\Bigg|\sum_{\substack{\frac{x}{Q|m|^2}<|n|^2\leqslant\frac{x}{|m|^2}\\ \gamma_1\leqslant\operatorname{Arg}(mn)<\gamma_2}} b_n g(mn)\Bigg| \leqslant D_0 + \sum_{1\leqslant|h|\leqslant H}\frac{1}{|h|}D_h + E, \tag{2.22}$$

where D_h is defined for all $h \geqslant 0$ by

$$D_h = \sum_{\frac{M}{Q}<|m|^2\leqslant M}\Bigg|\sum_{\frac{x}{Q|m|^2}<|n|^2\leqslant\frac{x}{|m|^2}} b_n\,\mathrm{e}(h\operatorname{Arg}(n))g(mn)\Bigg|,$$

and E is defined by

$$E = \sum_{\frac{M}{Q}<|m|^2\leqslant M}\sum_{\frac{x}{Q|m|^2}<|n|^2\leqslant\frac{x}{|m|^2}}\frac{1}{2H+2}\Big(\kappa_H(\operatorname{Arg}(mn)-\gamma_1)+\kappa_H(\operatorname{Arg}(mn)-\gamma_2)\Big).$$

In order to deal with D_h, we use the same considerations as in the first proof of Lemma 2.14. We get that D_h can be bounded by

$$\int_{-1/2}^{1/2}\min(x,|\sin\pi\xi|^{-1})\sum_{\frac{M}{Q}<|m|^2\leqslant M}\Bigg|\sum_{\frac{x}{QM}<|n|^2\leqslant\frac{xQ}{M}} b_n\,\mathrm{e}(h\operatorname{Arg}(n))\,\mathrm{e}(|n|^2\xi)g(mn)\Bigg|\,\mathrm{d}\xi,$$

and estimate (2.16) with $b_n\, \mathrm{e}(h \operatorname{Arg}(n))\, \mathrm{e}(|n|^2\xi)$ instead of b_n implies

$$D_h \ll_{q,\alpha} (\log x) x^{1-\hat{\sigma}''_q(\alpha)}. \tag{2.23}$$

As to obtain the desired result, we have to bound the error term E. First, we can extend the summation over m and n to a bigger domain since we are summing up only positive numbers (note, that κ_H is the positive Fejer kernel). Thus we get

$$E \leqslant \sum_{1 \leqslant |m|^2 \leqslant M} \sum_{1 \leqslant |n|^2 \leqslant \frac{xQ}{M}} \frac{1}{2H+2} \sum_{|h| \leqslant H} \left(1 - \frac{|h|}{H+1}\right) \cdot \Big(\mathrm{e}\big(h(\operatorname{Arg}(mn) - \gamma_1)\big) + \mathrm{e}\big(h(\operatorname{Arg}(mn) - \gamma_2)\big) \Big).$$

Now we can use again that $\operatorname{Arg}(mn) = \operatorname{Arg}(m) + \operatorname{Arg}(n) \bmod 1$. We obtain

$$\begin{aligned} E &\leqslant \frac{2}{2H+2} \sum_{|h| \leqslant H} \left| \sum_{1 \leqslant |m|^2 \leqslant M} \mathrm{e}(h \operatorname{Arg}(m) \sum_{1 \leqslant |n|^2 \leqslant \frac{xQ}{M}} \mathrm{e}(h \operatorname{Arg}(n)) \right| \\ &\leqslant \frac{4}{2H+2} \sum_{0 \leqslant h \leqslant H} \left| E_{h,M} \cdot E_{h,\frac{xQ}{M}} \right|, \end{aligned}$$

where $E_{h,K}$ is defined for all $h \geqslant 0$ and $K \geqslant 1$ by

$$E_{h,K} = \sum_{1 \leqslant |m|^2 \leqslant K} \mathrm{e}(h \operatorname{Arg}(m)).$$

We trivially have $E_{0,K} \ll K$. If $h > 0$, we use the Koksma-Hlawka inequality (see [DT97, Theorem 1.14]) in order to obtain

$$|E_{h,K}| \leqslant N_K (V(g_h^1) + V(g_h^2)) D^*_{N_K},$$

where $V(g_h^1)$ is the total variation (in the interval $[0,1]$) of the function $g_h^1(x) = \cos(2\pi h x)$, $V(g_h^2)$ is the total variation of the function $g_h^2(x) = \sin(2\pi h x)$, the number N_K is given by $N_K = \#\{z \in \mathbb{Z}[i] : |z|^2 \leqslant K\}$ and $D^*_{N_K}$ denotes the star discrepancy

$$D^*_{N_K} = \sup_{y \in [0,1]} \left| \frac{\#\{|z|^2 \leqslant K : 0 \leqslant \operatorname{Arg}(z) < y\}}{N_K} - y \right|.$$

We have $N_K = \pi K + O(K^{1/2})$. Geometric considerations show that $D^*_{N_K} \ll K^{-1/2}$. Since $V(g_h^1) = V(g_h^2) = 4h$, we obtain

$$|E_{h,K}| \ll h K^{1/2}$$

for $h \geqslant 1$. Hence, we get

$$E \ll \frac{1}{2H+2} \left(x + \sum_{1 \leqslant h \leqslant H} h^2 x^{1/2} \right) \ll \frac{x}{H} + H^2 x^{1/2}.$$

Choosing $H = x^{1/6}$, we finally obtain that $E \ll x^{5/6}$. This estimate together with (2.22) and (2.23) yields the desired result. □

Lemma 2.15. *Let $z_n \in \mathbb{C}$ with $n \in \mathbb{Z}[i]$ and A, B, $R \in \mathbb{R}$ with $1 < A < B$ and $R > 1$. Then we have*

$$\left| \sum_{A<|n|\leqslant B} z_n \right|^2 \leqslant C_1 \frac{(B - A + R + \min(A,R))(B + \max(A,R))}{R^2} \cdot \sum_{|r|\leqslant 2R} \frac{w(r)}{(2R+1)^2} \sum_{A<|n|,|n+r|\leqslant B} z_{n+r}\overline{z_n}, \tag{2.24}$$

where $C_1 = \frac{16}{9}\pi(1+\sqrt{2})$ and

$$w(r) = \#\left\{(r_1,r_2), 0 \leqslant |r_1|, |r_2| \leqslant R : r_1 - r_2 = r\right\} \leqslant (2R+1)(2R+1-|r|).$$

Remark 2.16. This result is a van der Corput-type inequality in $\mathbb{Z}[i]$ and a proof can be found in [DRS08, Lemma 3.4]. Since there is a small mistake in the statement of the cited lemma, we prove Lemma 2.15 for the sake of clearness. In particular, it is stated in [DRS08, Lemma 3.4] (under the same assumptions and with the same constant C_1) that

$$\left| \sum_{A<|n|\leqslant B} z_n \right|^2 \leqslant C_1 \frac{(B - A + R + \min(A,R))(B + \max(A,R))}{R^2} \cdot \sum_{|r|\leqslant 2R} \left(1 - \frac{|r|}{2R+1}\right) \sum_{A<|n|,|n+r|\leqslant B} z_{n+r}\overline{z_n}. \tag{2.25}$$

In what follows, we show that is false in general. We construct a family of examples for which the left-hand side of (2.25) is zero and the right-hand side tends to minus infinity.

Let $0 < \delta < 1/2$ and take $R = 3/2 - \delta$ and $A = 3/2$. Moreover, let the complex numbers a_z, $z \in \mathbb{Z}[i]$ be defined by

$$a_z = \begin{cases} 1, & \text{if } z = 3k+2, k \in \mathbb{N}, \\ -1, & \text{if } z = 3k+3, k \in \mathbb{N}, \\ 0, & \text{otherwise.} \end{cases}$$

If we take $B = 3K$ for some positive integer $K \geqslant 1$, then the left-hand side of (2.25) is zero while some tedious calculations show that the right-hand side is given by

$$C_1 \frac{(3K + 3/2 - 2\delta)(3K + 3/2)}{(3/2-\delta)^2} \left(-K\frac{1-2\delta}{2-\delta} + 4\frac{1-\delta}{2-\delta} \right).$$

If K goes to infinity, this value is not only negative but goes to minus infinity (in fact, for small values of δ it is negative if $K \geqslant 5$).

Nevertheless, if one takes the absolute value in (2.24) and uses the triangle inequality (which we actually do in (2.29) after changing an order of summation), we can apply the upper bound for $w(r)$.

Proof of Lemma 2.15. Set $z_n = 0$ for $|n| \notin (A, B]$ and put $T(R) = \#\{0 \leqslant |r| \leqslant R\}$. Then we have,

$$T(R) \sum_{n\in\mathbb{Z}[i]} z_n = \sum_{n} \sum_{0\leqslant|r|\leqslant R} z_{n+r} = \sum_{A-R<|n|\leqslant B+R} \sum_{0\leqslant|r|\leqslant R} z_{n+r}.$$

By the classical Gauss estimate for the number of lattice points in a disk we have

$$\#\{n \in \mathbb{Z}[i] : A - R < |n| \leqslant B + R\} \leqslant V,$$

where $V = (1+\sqrt{2})\pi(B - A + R + \min(A, R))(B + \max(A, R))$ (cf. [DRS08, p. 325]). Applying the Cauchy-Schwarz inequality yields

$$\begin{aligned} T(R)^2 \left|\sum_{n\in\mathbb{Z}[i]} z_n\right|^2 &\leqslant V \sum_{n\in\mathbb{Z}[i]} \left|\sum_{0\leqslant|r|\leqslant R} z_{n+r}\right|^2 \\ &\leqslant V \sum_{0\leqslant|r_1|,|r_2|\leqslant R} \sum_{n\in\mathbb{Z}[i]} z_{n+r_1}\overline{z_{n+r_2}} \\ &= V \sum_{0\leqslant|r|\leqslant 2R} w(r) \sum_{n\in\mathbb{Z}[i]} z_{n+r}\overline{z_n}, \end{aligned}$$

where $w(r)$ is equal to the number of pairs (r_1, r_2), such that $0 \leqslant |r_1|, |r_2| \leqslant R$ and $r_1 - r_2 = r$. Using $16T(R)^2/9 \geqslant R^2(2R + 1)^2$, we get the desired estimate. □

In order to employ Lemma 2.14, we need an estimate of the form (2.16). We set $f(n) = \alpha s_q(n)$ and define for every real positive number B the set $\varXi_B \subset \mathbb{Z}[i]$ by

$$\varXi_B := \{n \in \mathbb{Z}[i] : \max(|\Re(n)|, |\Im(n)|) \leqslant B^{1/2}\}.$$

Let

$$\mu := \left\lfloor \frac{\log M}{\log Q} \right\rfloor + 1 \qquad \text{and} \qquad \nu := \left\lfloor \frac{\log \frac{xQ}{M}}{\log Q} \right\rfloor + 1. \tag{2.26}$$

Furthermore, assume that ρ is an integers satisfying

$$1 \leqslant \rho \leqslant \nu/3. \tag{2.27}$$

We set

$$S := \sum_{\frac{M}{Q}<|m|^2\leqslant M} \left| \sum_{\frac{x}{QM}<|n|^2\leqslant\frac{xQ}{M}} b_n \,\mathrm{e}(f(mn)) \right|.$$

First of all, we trivially bound this double sum by

$$S \leqslant \sum_{m\in\varXi_{Q^\mu}} \left| \sum_{\frac{x}{QM}<|n|^2\leqslant\frac{xQ}{M}} b_n \,\mathrm{e}(f(mn)) \right|. \tag{2.28}$$

We want to remark, that this inconspicuous looking step is of great importance for the proof of Theorem 2.1 (compare with Remark 2.19). The Cauchy-Schwarz inequality implies

$$S^2 \ll Q^\mu \sum_{m\in\Xi_{Q^\mu}} \left| \sum_{\frac{x}{QM}<|n|^2\leqslant\frac{xQ}{M}} b_n \,\mathrm{e}(f(mn)) \right|^2 .$$

Set $A = (x/(QM))^{1/2}$, $B = (xQ/M)^{1/2}$, $R = (1/3)\,|q|^\rho$ and $z_n = b_n\,\mathrm{e}(f(mn))$. Since

$$\frac{1}{|q|^{2\rho}}\left((xQ/M)^{1/2} - (x/(QM))^{1/2} + \frac{2}{3}|q|^\rho\right)\left((xQ/M)^{1/2} + (x/(QM))^{1/2}\right) \ll Q^{\nu-\rho},$$

we get

$$\begin{aligned} S^2 \ll Q^{\mu+\nu-\rho} \sum_{m\in\Xi_{Q^\mu}} \sum_{|r|\leqslant\frac{2}{3}|q|^\rho} & \frac{w(r)}{(\frac{2}{3}|q|^\rho+1)^2} \\ \cdot \sum_{\substack{\frac{x}{QM}<|n|^2\leqslant\frac{xQ}{M} \\ \frac{x}{QM}<|n+r|^2\leqslant\frac{xQ}{M}}} & b_{n+r}\overline{b_n}\,\mathrm{e}\left(f(m(n+r)) - f(mn)\right). \end{aligned}$$

Changing the order of summation, taking the absolute value and using that $\{n \in \mathbb{Z}[i] : \frac{x}{QM} < |n|^2, |n+r|^2 \leqslant \frac{xQ}{M}\} \subseteq \Xi_{Q^\nu}$, we obtain

$$S^2 \ll Q^{\mu+\nu-\rho} \sum_{|r|\leqslant|q|^\rho}\left(1-\frac{|r|}{|q|^\rho}\right)\sum_{n\in\Xi_{Q^\nu}}\left|\sum_{m\in\Xi_{Q^\mu}} \mathrm{e}\left(f(m(n+r)) - f(mn)\right)\right|. \tag{2.29}$$

If we separate the case $r = 0$ and $r \neq 0$, we get (note, that $\rho \leqslant \nu/3$, see (2.27))

$$S^2 \ll Q^{2(\nu+\mu)-\rho} + Q^{\mu+\nu} \max_{1\leqslant|r|\leqslant|q|^\rho} \sum_{n\in\Xi_{Q^\nu}}\left|\sum_{m\in\Xi_{Q^\mu}} \mathrm{e}(f(m(n+r)) - f(mn))\right|.$$

In a next step, we use the fact that we are now dealing with expressions of the form $f(m(n+r)) - f(mn)$. If r is small (in comparison to n), the higher placed digits of $m(n+r)$ and mn do not differ in "most" of the cases. In order to show this, we define a "truncated" sum-of-digits function (times the constant α), namely,

$$f_\lambda(z) = \alpha\sum_{j=0}^{\lambda-1}\varepsilon_j(z),$$

where $\varepsilon_j(z)$, $j \geqslant 0$ are the digits of z in the base-q representation. The advantage of this function is that it is periodic with period q^λ, i.e., for any $d, z \in \mathbb{Z}[i]$, we have

$$f_\lambda\left(z + dq^\lambda\right) = f_\lambda(z). \tag{2.30}$$

For a proof of this statement, see [DRS08, Proposition 4.1]. Next, we briefly recall that the addition in the Gaussian integers can be realized by an automaton (see [GKP98]). For our further explanations we restrict ourselves to base $q = -a + i$, the case $q = -a - i$ is similar. The addition automaton in base $q = -a + i$ is drawn in Figure 2.1. The digits of the sum are associated to a walk which finishes in one of the two accepting states [•]. Starting at node **P**, it performs addition by 1 and starting at node **R** it performs addition by $-a - i$. The labeling $j|k$ means that the automaton reads a digit j and has k as output.

The following result ("carry lemma") allows us to replace the sum-of-digits function by its truncated version.

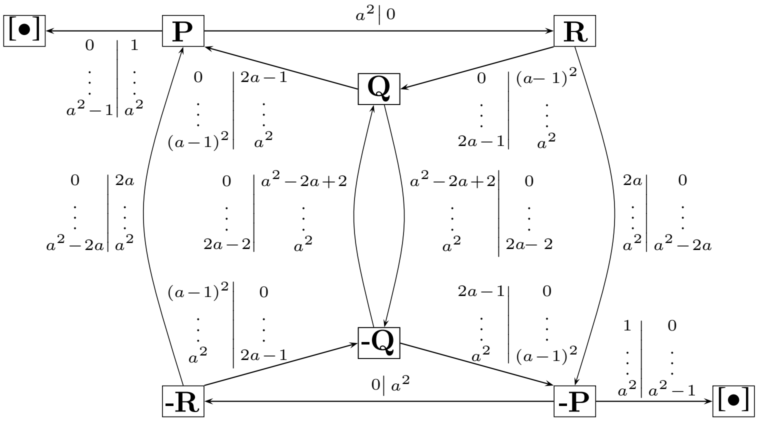

[•]
P
R
Q
-Q
-R
-P
[•]
$a^2 | 0$
$0 \ldots a^2-1 \,|\, 1 \ldots a^2$
$0 \ldots (a-1)^2 \,|\, 2a-1 \ldots a^2$
$0 \ldots 2a-1 \,|\, (a-1)^2 \ldots a^2$
$0 \ldots a^2-2a \,|\, 2a \ldots a^2$
$0 \ldots 2a-2 \,|\, a^2-2a+2 \ldots a^2$
$a^2-2a+2 \ldots a^2 \,|\, 0 \ldots 2a-2$
$2a \ldots a^2 \,|\, 0 \ldots a^2-2a$
$(a-1)^2 \ldots a^2 \,|\, 0 \ldots 2a-1$
$2a-1 \ldots a^2 \,|\, 0 \ldots (a-1)^2$
$1 \ldots a^2 \,|\, 0 \ldots a^2-1$
$0 | a^2$

Figure 2.1: Addition automaton in base $q = -a + i$

Lemma 2.17. *For all integers $\mu > 0$, $\nu > 0$ and $0 \leqslant \rho \leqslant \nu/3$ and $r \in \mathbb{Z}[i]$ with $|r|^2 < Q^\rho$, we denote by $E(r, \mu, \nu, \rho)$ the set of pairs of Gaussian integers (m, n) such that $m \in \Xi_{Q^\mu}$, $n \in \Xi_{Q^\nu}$ and*

$$f(m(n+r)) - f(mn) \neq f_{\mu+2\rho}(m(n+r)) - f_{\mu+2\rho}(mn).$$

Then we have for any $\varepsilon > 0$,

$$\#E(r, \mu, \nu, \rho) \ll_\varepsilon Q^{(\mu+\nu)(1+\varepsilon)-\iota\rho},$$

where $0 < \iota < 1$ is a constant only depending on q.

Remark 2.18. This lemma is similar to [DRS08, Lemma 4.2]. Since there seem to be some mistakes at the end of the proof of [DRS08, Lemma 4.2], we give a new proof of Lemma 2.17. Note, that our statement is a little bit weaker than the statement of Drmota et al., but it suffices for our purposes and it would also suffice for their further reasoning.

Proof of Lemma 2.17. It follows from Proposition A.5 that there exists a constant c (only depending on q) such that for all $m \in \varXi_{Q^\mu}$, $n \in \varXi_{Q^\nu}$ and $|r|^2 < Q^\rho$ the number mn has $\leqslant \mu + \nu + c$ significant digits, i.e., $mn \in \mathcal{F}_{\mu+\nu+c}$ (see Section 2.2 for the definition of $\mathcal{F}_\lambda$) and mr has $\leqslant \mu + \rho + c$ significant digits ($mr \in \mathcal{F}_{\mu+\rho+c}$). We can assume that $\rho > c$ (the statement is trivial in the converse case). The main idea of the proof of this lemma is the fact that every Gaussian integer z with $\leqslant \mu + \nu + c$ digits can be uniquely written as

$$z = w_0(z) + q^{\mu+\rho+c} w_1(z) + q^{\mu+2\rho} w_2(z),$$

where $w_0(z) \in \mathcal{F}_{\mu+\rho+c}$, $w_1(z) \in \mathcal{F}_{\rho-c}$ and $w_2(z) \in \mathcal{F}_{\nu+c-2\rho}$. We set

$$T = \{t \in \mathcal{F}_{\rho-c} : \ t = w_1(mn) \text{ for some } (m,n) \in E(r,\mu,\nu,\rho)\},$$

and

$$N_t = \{(m,n) \in E(r,\mu,\nu,\rho) : w_1(mn) = t\}.$$

In what follows we show that

$$\# T \ll Q^{\iota'\rho}, \tag{2.31}$$

where ι' is a positive constant satisfying $\iota' < 1$ and

$$\# N_t \ll_\varepsilon Q^{(\mu+\nu)(1+\varepsilon)-\rho} \tag{2.32}$$

for any $\varepsilon > 0$. The statement of the lemma is then a direct consequence of these two estimates since

$$\# E(r,\mu,\nu,\rho) = \sum_{t\in T} \# N_t \ll_\varepsilon \sum_{t \in T} Q^{(\mu+\nu)(1+\varepsilon)-\rho} \ll_\varepsilon Q^{(\mu+\nu)(1+\varepsilon)-(1-\iota')\rho}.$$

Indeed, setting $\iota = 1 - \iota' > 0$ proves the desired result.

First we show (2.31), i.e., we bound the number of possible numbers t which lead to a carry propagation. Let $(m,n) \in E(r,\mu,\nu,\rho)$. We can write

$$mn + mr = (w_0(mn) + mr) + \left(q^{\mu+\rho+c} w_1(mn) + q^{\mu+2\rho} w_2(mn)\right).$$

If we consider only the first part of the sum, we obtain that

$$w_0(mn) + mr = \widetilde{w} + q^{\mu+\rho+c}(x+iy),$$

where $\widetilde{w} \in \mathcal{F}_{\mu+\rho+c}$ is some Gaussian integer depending on m, n and r and $x, y \in \mathbb{Z}$ with $x = O(1)$ and $y = O(1)$ (this is a consequence of Proposition A.5). Next we claim that

$$f(w_1(mn) + x + iy) - f(w_1(mn)) \neq f_{\rho-c}(w_1(mn) + x + iy) - f_{\rho-c}(w_1(mn)). \tag{2.33}$$

First note that $f(w_1(mn)) = f_{\rho-c}(w_1(mn))$. If we assume equality in (2.33), we conclude that $w_1(mn) + x + iy$ has at most $\rho - c$ digits. It follows that

$$\begin{aligned} f\left(mn+mr\right) - f\left(mn\right) &= f\left(\widetilde{w} + q^{\mu+\rho+c}\left(x+iy+w_1(mn)\right) + q^{\mu+2\rho}w_2(mn)\right) \\ &\quad - f\left(w_0(mn) + q^{\mu+\rho+c}w_1(mn) + q^{\mu+2\rho}w_2(mn)\right) \\ &= f\left(\widetilde{w} + q^{\mu+\rho+c}\left(x+iy+w_1(mn)\right)\right) \\ &\quad - f\left(w_0(mn) + q^{\mu+\rho+c}w_1(mn)\right). \end{aligned}$$

The integers in the last expression have at most $\mu + 2\rho$ digits and we obtain

$$\begin{aligned} f\left(mn+mr\right) - f\left(mn\right) &= f_{\mu+2\rho}\left(\widetilde{w} + q^{\mu+\rho+c}\left(x+iy+w_1(mn)\right)\right) \\ &\quad - f_{\mu+2\rho}\left(w_0(mn) + q^{\mu+\rho+c}w_1(mn)\right) \\ &= f_{\mu+2\rho}\left(mn+mr\right) - f_{\mu+2\rho}\left(mn\right). \end{aligned}$$

This contradicts $(m,n) \in E(r,\mu,\nu,\rho)$ and (2.33) holds true, indeed. From now on we assume that $x+iy = -y(-a-i)+(x-ay)$ with $y > 0$ and $x < ay$ (all other cases are similar). Using an idea developed in [GKP98, Proposition 2.4], the left-hand side of (2.33) can be written as

$$\begin{aligned} &f(w_1(mn)+x+iy) - f(w_1(mn)) = \\ &\quad f\left(w_1(mn)+x+iy\right) - f\left(w_1(mn)+(-a-i)+x+iy\right) \\ &\quad +f\left(w_1(mn)+(-a-i)+x+iy\right) - f\left(w_1(mn)+2(-a-i)+x+iy\right) + \dots \\ &\quad +f\left(w_1(mn)+(y-1)(-a-i)+x+iy\right) - f\left(w_1(mn)+y(-a-i)+x+iy\right) \\ &\quad +f\left(w_1(mn)+y(-a-i)+x+iy\right) - f\left(w_1(mn)+y(-a-i)+x+iy+1\right) \\ &\quad +f\left(w_1(mn)+y(-a-i)+x+iy+1\right) - f\left(w_1(mn)+y(-a-i)+x+iy+2\right) + \dots \\ &\quad +f\left(w_1(nm)+y(-a-i)+x+iy+(-x+ay-1)\right) - f\left(w_1(mn)\right). \end{aligned}$$

The number of differences is $O(1)$. Since one can write the right-hand side of (2.33) in the same way (replacing f by $f_{\rho-c}$), we deduce that (2.33) can be only fulfilled if one of the differences is not equal to the corresponding one with f replaced by $f_{\rho-c}$. Furthermore each of them is of the form $f(x) - f(x+u)$, where u is either $-a-i$ or 1. Thus we are interested in the number of cases where the addition $x \mapsto x+u$ gives rise to a carry propagation.

Recall, that the addition in the Gaussian integers can be realized by an automaton. There is a carry propagation if the addition automaton does not end in one of the two accepting states after reading the first $\rho - c$ digits. The same reasoning as in [DRS08, bottom of page 329] shows (cf. [GL99, Proposition 1]) that the number of possible inputs such that this occurs is bounded by $O\left(|\xi|^{\rho}\right)$, where $|\xi| = Q^{\iota'}$ with $\iota' < 1$. It finally follows that $\# T \ll Q^{\iota'\rho}$. In order to complete the proof, it remains to show that (2.32) holds true. We have

$$\#N_t = \sum_{\substack{a\in\mathcal{F}_{\mu+\nu+c} \\ w_1(a)=t}} \#\{(m,n) \in E(r,\mu,\nu,\rho) : mn = a\} \leqslant \sum_{\substack{a\in\mathcal{F}_{\mu+\nu+c} \\ w_1(a)=t}} \tau(a),$$

where $\tau(a)$ denotes the number of divisors of a. But since $\tau(a) \ll_\varepsilon |a|^{2\varepsilon}$ and since the number of Gaussian integers a satisfying $a \in \mathcal{F}_{\mu+\nu+c}$ and $w_1(a) = t$ is $Q^{\mu+\nu+2c-\rho}$ (note, that $\rho - c$ digits are fixed), we are done. □

We set

$$\lambda = \mu + 2\rho. \tag{2.34}$$

Replacing f by f_λ gives a total error of $O(Q^{2(\mu+\nu)(1+\varepsilon)-\iota\rho})$ to $|S|^2$. Thus, we have

$$S^2 \ll_\varepsilon Q^{2(\mu+\nu)(1+\varepsilon)-\iota\rho} + Q^{\mu+\nu} \max_{1 \leqslant |r| \leqslant |q|^\rho} |S_2(r,\mu,\nu,\rho)|, \tag{2.35}$$

where

$$S_2(r,\mu,\nu,\rho) = \sum_{n \in \Xi_{Q^\nu}} \left| \sum_{m \in \Xi_{Q^\mu}} \mathrm{e}(f_\lambda(m(n+r)) - f_\lambda(mn)) \right|,$$

and $\iota > 0$. The inverse of the Fourier transform (defined in Section 2.2) is given by

$$\mathrm{e}\left(f_\lambda(z)\right) = \sum_{h \in \mathcal{F}_\lambda} F_\lambda(h,\alpha)\, \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{zh}{q^\lambda}\right)\right).$$

Hence, we obtain

$$\begin{aligned}\sum_{m \in \Xi_{Q^\mu}} &\mathrm{e}(f_\lambda(m(n+r)) - f_\lambda(mn)) \\ &= \sum_{h,k \in \mathcal{F}_\lambda} F_\lambda(h,\alpha)\overline{F_\lambda(-k,\alpha)} \sum_{m \in \Xi_{Q^\mu}} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{hm(n+r)+kmn}{q^\lambda}\right)\right).\end{aligned}$$

Remark 2.19. At this point we use the fact that the domain of summation over m is equal to Ξ_{Q^μ} (compare with (2.28)). In contrast to [DRS08], we do not have to deal with the classical circle problem (cf. [DRS08, Lemma 5.1 and the subsequent remark (p.332)]). Indeed, calculating two times (independently) a geometric series, we obtain

$$\sum_{m \in \Xi_M} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left((r+is)m\right)\right) \ll \min\left(M^{1/2}, \frac{1}{\|r\|}\right) \cdot \min\left(M^{1/2}, \frac{1}{\|s\|}\right).$$

If we set $\tau(z) = (\max(\|\Re(z)\|, \|\Im(z)\|))^{-1}$ and $\xi(z) = (\|\Re(z)\| \cdot \|\Im(z)\|)^{-1}$, this can be also written in the form

$$\sum_{m \in \Xi_M} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(zm\right)\right) \ll \min\left(M, M^{1/2}\tau(z), \xi(z)\right).$$

In the end, the additional term $\xi(z)$ on the right-hand side of the last inequality makes it possible for us to show Theorem 2.1 in its full generality. In particular, we will use that $\xi(z) = \min\left(M, M^{1/2}\tau(z), \xi(z)\right)$ if z is a complex number such that $\|\Re(z)\|$ and $\|\Im(z)\|$ are not too small (see Lemma 2.20 and Lemma 2.21).

The last remark implies

$$S_2 \ll \sum_{h,k\in\mathcal{F}_\lambda} |F_\lambda(h,\alpha)F_\lambda(-k,\alpha)| \cdot \sum_{n\in\Xi_{Q^\nu}} \min\left(Q^\mu, Q^{\mu/2}\tau\left(\frac{(h+k)n+hr}{q^\lambda}\right), \xi\left(\frac{(h+k)n+hr}{q^\lambda}\right)\right).$$

If $z\in\mathbb{C}$ we denote by $\lfloor z\rfloor$ the Gaussian integer w_1+iw_2, such that $w_1=\lfloor\Re(z)\rfloor$ and $w_2=\lfloor\Im(z)\rfloor$. Recall that R_m denotes a complete residue system modulo m (see Section 2.2). The next lemma is a refinement of [GT00, Lemma 2.6], see also Lemma 3.14 and Remark 3.15.

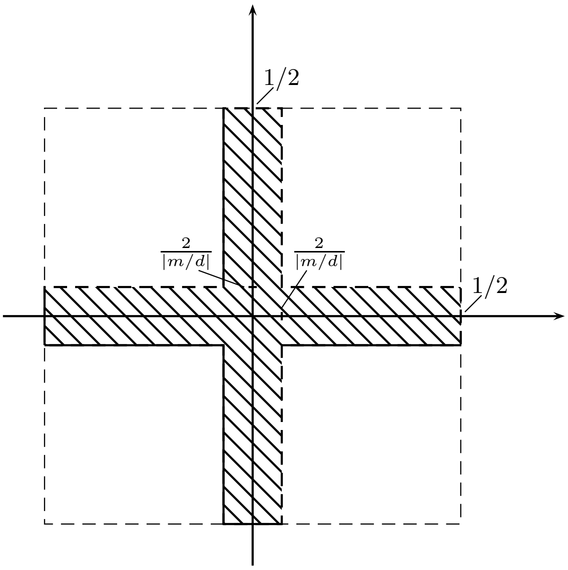


Figure 2.2: The domain R^+ in $\mathbb{R}^2$

Lemma 2.20. *Let $m,b,c\in\mathbb{Z}[i]$ with $m\neq 0$ and $M\geqslant 0$. Set $d=(c,m)$. Then*

$$\sum_{n\in R_m} \min\left(M, M^{1/2}\,\tau\left(\frac{cn+b}{m}\right), \xi\left(\frac{cn+b}{m}\right)\right) \ll |d|^2 \sum_{n\in\mathcal{R}^+} \min\left(M, M^{1/2}\,\tau\left(\frac{n}{m/d}+t\right), \xi\left(\frac{n}{m/d}+t\right)\right) + |m|^2(\log|m|)^2,$$

where $t=(b-\lfloor b/d\rfloor d)/m$, $\mathcal{R}^+=\mathbb{Z}[i]\cap\{(\alpha+i\beta)m/d : (\alpha,\beta)\in R^+\}$ and R^+ equals

$$\left[-\frac{1}{2},\frac{1}{2}\right)^2 \cap \left(\left\{(\alpha,\beta) : -\frac{2|d|}{|m|}\leqslant\alpha<\frac{2|d|}{|m|}\right\}\cup\left\{(\alpha,\beta) : -\frac{2|d|}{|m|}\leqslant\beta<\frac{2|d|}{|m|}\right\}\right).$$

Proof. Let us denote the considered sum with T. If $|d|=|m|$ then

$$\tau\left(\frac{cn+b}{m}\right)=\tau\left(\frac{b}{d}\right)=\tau\left(\frac{nd}{m}+\frac{b-\lfloor\frac{b}{d}\rfloor d}{m}\right)$$

for all $n \in \mathbb{Z}[i]$ (note, that $\tau(\varepsilon z) = \tau(z)$ for every unit $\varepsilon \in \{\pm 1, \pm i\}$ and for every $z \in \mathbb{C}$). The same holds for ξ and the result of the lemma follows trivially. Let $|d| \neq |m|$, then we have $1 \leqslant |d| \leqslant |m|/\sqrt{2}$. Set

$$c' = c/d \quad \text{and} \quad m' = m/d.$$

Note that $P = \mathbb{Z}[i] \cap \{(\alpha + i\beta)m' : -1/2 \leqslant \alpha, \beta < 1/2\}$ is a complete residue system modulo m' and that the summands of T are periodic with period m'. We can completely tessellate (a proper chosen residue system) R_m with at most $O(|d|^2)$ translates of P. Since $c'n + \lfloor b/d \rfloor$ covers also all residue classes modulo m', we can write

$$T \ll |d|^2 \sum_{n \in P} \min\left(M, M^{1/2}\, \tau\Big(\frac{n}{m'} + t\Big), \xi\left(\frac{n}{m'} + t\right)\right),$$

where $t = (b - \lfloor b/d \rfloor\, d)/m$. In order to prove the desired result, it remains to show that

$$|d|^2 \sum_{(\alpha,\beta) \in \hat{R}} \min\Big(M, M^{1/2}\, \tau(\alpha + i\beta + t)\, , \xi\,(\alpha + i\beta + t)\Big) \ll |m|^2 (\log |m|)^2,$$

where $\hat{R}$ is defined by $\hat{R} = \{(\alpha, \beta) \in [-1/2, 1/2)^2 : (\alpha + i\beta)m' \in P \setminus \mathcal{R}^+\}$. In a first step, we tessellate the domain $\hat{R}$ with small squares of side length $2/|m'|$. Therefore, we set

$$\hat{R}_{h,k} := \left[\frac{2h}{|m'|}, \frac{2(h+1)}{|m'|}\right) \times \left[\frac{2k}{|m'|}, \frac{2(k+1)}{|m'|}\right),$$

and

$$\hat{\mathcal{R}}_{h,k} := \hat{R}_{h,k} \cup \hat{R}_{-h-1,k} \cup \hat{R}_{-h-1,-k-1} \cup \hat{R}_{h,-k-1}.$$

Using this notation, we can write

$$\begin{aligned}
&|d|^2 \sum_{(\alpha,\beta) \in \hat{R}} \min\Big(M, M^{1/2}\, \tau(\alpha + i\beta + t)\, , \xi\,(\alpha + i\beta + t)\Big) \\
&\quad = |d|^2 \sum_{h=1}^{\lfloor |m'| \rfloor/4} \sum_{k=1}^{\lfloor |m'| \rfloor/4} \sum_{(\alpha,\beta) \in \hat{R} \cap \hat{\mathcal{R}}_{h,k}} \min\Big(M, M^{1/2}\, \tau(\alpha + i\beta + t)\, , \xi\,(\alpha + i\beta + t)\Big).
\end{aligned}$$

If $(\alpha, \beta) \in \hat{\mathcal{R}}_{h,k}$, then

$$\xi\,(\alpha + i\beta + t) \ll \frac{1}{h/|m'| \cdot k/|m'|}$$

for all $h, k \geqslant 1$ (note, that $|t| \leqslant \sqrt{2}/|m'|$). Furthermore, we have $\#(\hat{R} \cap \hat{\mathcal{R}}_{h,k}) \ll 1$ and we get

$$\begin{aligned}
&|d|^2 \sum_{h=1}^{\lfloor |m'| \rfloor/4} \sum_{k=1}^{\lfloor |m'| \rfloor/4} \sum_{(\alpha,\beta) \in \hat{R} \cap \hat{\mathcal{R}}_{h,k}} \min\Big(M, M^{1/2}\, \tau(\alpha + i\beta + t)\, , \xi\,(\alpha + i\beta + t)\Big) \\
&\quad \ll |d|^2 \left(\sum_{h=1}^{\lfloor |m'| \rfloor/4} \frac{|m'|}{h}\right)^2 \ll |d|^2 |m'|^2 (\log |m|)^2 = |m|^2 (\log |m|)^2.
\end{aligned}$$

As noticed above, this finally shows the desired result. □

In order to apply Lemma 2.20, we tessellate Ξ_{Q^ν} by translates of an appropriate complete residue system R_{q^λ}. If $\lambda \geqslant \nu$, the number of translates is $\ll 1$, otherwise it is $\ll Q^{\nu-\lambda}$. We obtain

$$S_2 \ll (1+Q^{\nu-\lambda})\left(S_3^{(1)}+S_3^{(2)}\right), \tag{2.36}$$

where

$$\begin{aligned} S_3^{(1)} := \sum_{d \mid q^\lambda} \sum_{\substack{h,k\in\mathcal{F}_\lambda\\ (h+k,q^\lambda)=d}} &|F_\lambda(h,\alpha)F_\lambda(-k,\alpha)| \\ &\cdot |d|^2 \sum_{n\in\mathcal{R}^+} \min\left(Q^\mu, Q^{\mu/2}\,\tau\left(\frac{nd}{q^\lambda}+\frac{hr-\left\lfloor\frac{hr}{d}\right\rfloor d}{q^\lambda}\right), \xi\left(\frac{nd}{q^\lambda}+\frac{hr-\left\lfloor\frac{hr}{d}\right\rfloor d}{q^\lambda}\right)\right), \end{aligned}$$

and

$$S_3^{(2)} := (\log Q^\lambda)^2 Q^\lambda \sum_{h,k\in\mathcal{F}_\lambda} |F_\lambda(h,\alpha)F_\lambda(-k,\alpha)|.$$

The sum $S_3^{(2)}$ can be bounded above with the help of Lemma 2.7 and Lemma 2.8 (see Remark 2.9). We get

$$S_3^{(2)} \ll (\log Q^\lambda)^2 Q^{(1+2\eta)\lambda}. \tag{2.37}$$

Next, we treat the expression $S_3^{(1)}$. First, we replace $(h+k,q^\lambda)=d$ by the less restrictive condition $h+k\equiv 0 \bmod d$. We can separate this condition into $h\equiv b \bmod d$ and $k\equiv -b \bmod d$, where b covers all residue classes modulo d. We obtain

$$\begin{aligned} S_3^{(1)} \leqslant \sum_{d\mid q^\lambda} |d|^2 \sum_{b\in R_d}\sum_{n\in\mathcal{R}^+} &\left(\sum_{\substack{h\in\mathcal{F}_\lambda\\ h\equiv b \bmod d}} |F_\lambda(h,\alpha)|\right)^2 \\ &\cdot \min\left(Q^\mu, Q^{\mu/2}\,\tau\left(\frac{nd}{q^\lambda}+\frac{br-\left\lfloor\frac{br}{d}\right\rfloor d}{q^\lambda}\right), \xi\left(\frac{nd}{q^\lambda}+\frac{br-\left\lfloor\frac{br}{d}\right\rfloor d}{q^\lambda}\right)\right). \end{aligned}$$

Hence we get (see Remark 2.9)

$$\begin{aligned} S_3^{(1)} \ll Q^{2\eta\lambda} \sum_{d\mid q^\lambda} |d|^{2-4\eta} &\sum_{b\in R_d} |F_{\nu_q(d)}(b,\alpha)|^2 \\ &\cdot \sum_{n\in\mathcal{R}^+} \min\left(Q^\mu, Q^{\mu/2}\,\tau\left(\frac{nd}{q^\lambda}+\frac{br-\left\lfloor\frac{br}{d}\right\rfloor d}{q^\lambda}\right), \xi\left(\frac{nd}{q^\lambda}+\frac{br-\left\lfloor\frac{br}{d}\right\rfloor d}{q^\lambda}\right)\right), \end{aligned}$$

where $\nu_q(d)$ denotes the unique integer k such that $q^k \mid d$ but $q^{k+1} \nmid d$. Next, we use the L^∞-norm estimate of the Fourier transform. Note that Drmota et al. applied this estimate at a different place in their proof (which forced them to use a new parameter λ, see [DRS08, Section 7]). But similar to [MMR], we can simplify our further considerations if we use it right now. The notation of this part of the proof is also inspired by [MMR]. Lemma 2.5 implies

$$S_3^{(1)} \ll Q^{2\eta\lambda} \sum_{d \mid q^\lambda} |d|^{2-4\eta} Q^{-2c_Q \left\|(a^2+2a+2)\alpha\right\|^2 \nu_q(d)}$$
$$\cdot \sum_{n \in \mathcal{R}^+} \sum_{b \in R_d} \min\left(Q^\mu, Q^{\mu/2}\, \tau\left(\frac{nd}{q^\lambda} + \frac{br - \left\lfloor \frac{br}{d} \right\rfloor d}{q^\lambda}\right), \xi\left(\frac{nd}{q^\lambda} + \frac{br - \left\lfloor \frac{br}{d} \right\rfloor d}{q^\lambda}\right)\right),$$

where c_Q is defined in Remark 2.6. We use once more the fact that the term $\xi(z)$ is small (in comparison to the other considered terms) in "most" of the cases (cf. Remark 2.19).

Lemma 2.21. *Let $m, d, r \in \mathbb{Z}[i]$ with $m \neq 0$ and $d \mid m$. Furthermore, let $M \geqslant 0$ and $\mathcal{R}^+$ be defined as in Lemma 2.20. Set $e = (d, r)$. Then*

$$\sum_{n \in \mathcal{R}^+} \sum_{b \in R_d} \min\left(M, M^{1/2}\, \tau\left(\frac{nd}{m} + \frac{br - \left\lfloor \frac{br}{d} \right\rfloor d}{m}\right), \xi\left(\frac{nd}{m} + \frac{br - \left\lfloor \frac{br}{d} \right\rfloor d}{m}\right)\right)$$
$$\ll M|e|^2 + M^{1/2}|e||m| \log|m| + |m|^2 (\log|m|)^2.$$

Proof. Set $\tilde{r} = r/e$, $\tilde{d} = d/e$ and $\tilde{m} = m/e$. Furthermore, let R^+ be defined as in Lemma 2.20 (see Figure 2.2) and set $t(b) = \left(b\tilde{r} - \left\lfloor \frac{b\tilde{r}}{\tilde{d}} \right\rfloor \tilde{d}\right)/\tilde{m}$. Then we have that

$$\min\left(M, M^{1/2}\, \tau(\alpha + i\beta + t(b)), \xi(\alpha + i\beta + t(b))\right)$$

is periodic with period $\tilde{d}$ in b. We set $\hat{R} = \{(\alpha, \beta) : (\alpha + i\beta)m/d \in \mathcal{R}^+\}$ and denote the considered sum in the statement of Lemma 2.21 by T. Then we can write

$$T = |e|^2 \sum_{(\alpha,\beta) \in \hat{R},\, b \in R_{\tilde{d}}} \min\left(M, M^{1/2}\, \tau(\alpha + i\beta + t(b)), \xi(\alpha + i\beta + t(b))\right).$$

Next, we tessellate a neighborhood of the domain R^+ with small squares of side length $1/|\tilde{m}|$. Following the notion of Lemma 2.20, we set

$$\hat{R}_{h,k} := \left[\frac{h}{|\tilde{m}|}, \frac{h+1}{|\tilde{m}|}\right) \times \left[\frac{k}{|\tilde{m}|}, \frac{k+1}{|\tilde{m}|}\right),$$

and

$$\hat{\mathcal{R}}_{h,k} := \hat{R}_{h,k} \cup \hat{R}_{-h-1,k} \cup \hat{R}_{-h-1,-k-1} \cup \hat{R}_{h,-k-1}.$$

We obtain

$$T \ll |e|^2 \sum_{h=0}^{\lfloor|\tilde{m}|\rfloor} \sum_{k=0}^{4\lfloor|\tilde{d}|\rfloor}$$
$$\cdot \sum_{\substack{(\alpha,\beta) \in \hat{R},\, b \in R_{\tilde{d}} \\ \alpha + i\beta + t(b) \in \hat{\mathcal{R}}_{h,k} \cup \hat{\mathcal{R}}_{k,h}}} \min\left(M, M^{1/2}\, \tau(\alpha + i\beta + t(b)), \xi\left(\alpha + i\beta + t(b)\right)\right). \tag{2.38}$$

Next, we show that the inner sum in (2.38) has $\ll 1$ summands. This follows from the fact the sets $\{\alpha + i\beta : (\alpha, \beta) \in \hat{R}\}$ and $\{t(b) : b \in R_{\tilde{d}}\}$ are well-spaced (that is, $|\alpha_1 + i\beta_1 - (\alpha_2 + i\beta_2)| \geqslant |d|/|m| = |\tilde{d}|/|\tilde{m}|$ for all $(\alpha_1, \beta_1) \neq (\alpha_2, \beta_2)$ and $|t(b_1) - t(b_2)| \geqslant 1/|\tilde{m}|$ for all $b_1 \neq b_2$). If we fix h and k, then the number of possible pairs (α, β) such that there exists a number $b \in R_{\tilde{d}}$ with $\alpha + i\beta + t(b) \in \hat{R}_{h,k}$ is bounded above by some absolute constant (note, that $|t(b)| \leqslant \sqrt{2}|\tilde{d}|/|\tilde{m}|$). But for each fixed pair (α, β) there are $\leqslant 4$ numbers $b \in R_{\tilde{d}}$ with $\alpha + i\beta + t(b) \in \hat{R}_{h,k}$. Hence, we have indeed that

$$\#\{(\alpha, \beta) \in \hat{R},\, b \in R_{\tilde{d}} : \alpha + i\beta + t(b) \in \hat{\mathcal{R}}_{h,k} \cup \hat{\mathcal{R}}_{k,h}\} \ll 1.$$

In order to estimate the sum T, we distinguish between three different cases. If $z \in \hat{\mathcal{R}}_{0,0}$, we use

$$\min\left(M, M^{1/2}\,\tau(z)\,, \xi\,(z)\right) \leqslant M.$$

If $z \in \hat{\mathcal{R}}_{h,0}$ with $h > 0$, we have

$$\min\left(M, M^{1/2}\,\tau(z)\,, \xi\,(z)\right) \leqslant \frac{M^{1/2}}{h/|\tilde{m}|}.$$

A corresponding result holds for $z \in \hat{\mathcal{R}}_{0,k}$ with $k > 0$. If finally $z \in \hat{\mathcal{R}}_{h,k}$ with $h > 0$ and $k > 0$, we have

$$\min\left(M, M^{1/2}\,\tau(z)\,, \xi\,(z)\right) \leqslant \frac{1}{h/|\tilde{m}|} \cdot \frac{1}{k/|\tilde{m}|}.$$

Thus, we obtain

$$\begin{aligned} T &\ll |e|^2 \left(M + M^{1/2} \sum_{h=1}^{\lfloor |\tilde{m}| \rfloor} \frac{1}{h/|\tilde{m}|} + \left(\sum_{h=1}^{\lfloor |\tilde{m}| \rfloor} \frac{1}{h/|\tilde{m}|} \right)^2 \right) \\ &\ll |e|^2 M + |e|^2 M^{1/2} |\tilde{m}| \log |\tilde{m}| + |e|^2 |\tilde{m}|^2 (\log |\tilde{m}|)^2, \end{aligned}$$

which determines the proof of Lemma 2.21. □

Recall that we have $|(r, d)| \leqslant |r| \leqslant |q|^\rho$ and $\lambda = \mu + 2\rho$ (see (2.34) and (2.35)). Since

$$Q^\rho Q^\mu + Q^{\mu/2} Q^{(\lambda+\rho)/2} \log Q^\lambda + Q^\lambda (\log Q^\lambda)^2 \ll Q^\lambda (\log Q^\lambda)^2,$$

we obtain

$$\begin{aligned} S_3^{(1)} &\ll (\log Q^\lambda)^2 Q^{(1+2\eta)\lambda} \sum_{d | q^\lambda} |d|^{2-4\eta} Q^{-2c_Q \left\| (a^2+2a+2)\alpha \right\|^2 \nu_q(d)} \\ &\ll (\log Q^\lambda)^2 Q^{(1+2\eta)\lambda} \sum_{0 \leqslant \delta \leqslant \lambda} Q^{(1-2\eta-2c_Q \left\| (a^2+2a+2)\alpha \right\|^2)\delta} \sum_{\substack{k | q^{\lambda-\delta} \\ q \nmid k}} |k|^{2(1-2\eta)}. \end{aligned}$$

We have for every $k \mid q^{\lambda-\delta}$ with $q \nmid k$ that

$$|k| = |(k, q^{\lambda-\delta})| \leqslant |(k,q)|^{\lambda-\delta} \leqslant \left(\frac{|q|}{\sqrt{2}}\right)^{\lambda-\delta} = Q^{\frac{\lambda-\delta}{2}(1-\log_Q 2)}.$$

Recall, that $\tau(m)$ denotes the number of divisors of m. We get

$$\begin{aligned} S_3^{(1)} &\ll \tau(q^\lambda)(\log Q^\lambda)^2 Q^{(1+2\eta)\lambda} \sum_{0\leqslant\delta\leqslant\lambda} Q^{(1-2\eta-2c_Q\left\|(a^2+2a+2)\alpha\right\|^2)\delta+(\lambda-\delta)(1-2\eta)(1-\log_Q 2)} \\ &\ll \tau(q^\lambda)(\log Q^\lambda)^2 Q^{(2-(1-2\eta)\log_Q 2)\lambda} \sum_{0\leqslant\delta\leqslant\lambda} Q^{((1-2\eta)\log_Q 2-2c_Q\left\|(a^2+2a+2)\alpha\right\|^2)\delta}. \end{aligned}$$

We have (see Section 2.2 for the definition of η and c_Q)

$$\begin{aligned} (1-2\eta)\log_Q 2 - 2c_Q\left\|(a^2+2a+2)\alpha\right\|^2 &\geqslant (1-2\eta)\log_Q 2 - c_Q/2 \\ &= \frac{1}{\log Q}\left((1-2\eta)\log 2 - \frac{\pi^2}{54}\left(\frac{Q^2-1}{Q^4}\right)\right) \\ &\geqslant \frac{1}{\log Q}\left((1-2\eta)\log 2 - \frac{\pi^2}{54}\left(\frac{2^2-1}{2^4}\right)\right) \\ &> \frac{0.01428}{\log Q} > 0. \end{aligned} \tag{2.39}$$

Essentially, we use the fact that $\eta < 1/2$. If η were closer to $1/2$ (but still smaller than $1/2$), then we just would have to scale down c_Q by a constant factor (here we have $c_Q \leqslant \pi^2/(144\log Q) < 0.099$). We get

$$S_3^{(1)} \ll_q \tau(q^\lambda)(\log Q^\lambda)^2 Q^{(2-2c_Q\left\|(a^2+2a+2)\alpha\right\|^2)\lambda}, \tag{2.40}$$

and (cf. (2.36), (2.37) and (2.40))

$$S_2 \ll_q \tau(q^\lambda)(\log Q^\lambda)^2(1+Q^{\nu-\lambda})\left(Q^{(1+2\eta)\lambda} + Q^{(2-2c_Q\left\|(a^2+2a+2)\alpha\right\|^2)\lambda}\right).$$

Exactly the same way as in [MR09, Lemma 20], one can show that $\tau(q^\lambda) \leqslant \lambda^{\omega(q)}\tau(q)$, where $\omega(q)$ denotes the number of distinct prime divisors of q. Similar to (2.39), we see that $2-2c_Q\left\|(a^2+2a+2)\alpha\right\|^2 > 1+2\eta$ (actually, here we use again that $\eta < 1/2$). This allows us to write

$$\begin{aligned} S_2 &\ll_q \lambda^{\omega(q)+2}(1+Q^{\nu-\lambda})Q^{(2-2c_Q\left\|(a^2+2a+2)\alpha\right\|^2)\lambda} \\ &\ll_q (\mu+\nu)^{\omega(q)+2}\left(Q^{(2-2c_Q\left\|(a^2+2a+2)\alpha\right\|^2)\mu+4\rho} + Q^{(1-2c_Q\left\|(a^2+2a+2)\alpha\right\|^2)\mu+\nu+2\rho}\right). \end{aligned} \tag{2.41}$$

If the conditions

$$\frac{1}{\mu+\nu}\cdot\frac{3\rho}{2c_Q\left\|(a^2+2a+2)\alpha\right\|^2} \leqslant \frac{\mu}{\mu+\nu} \leqslant \frac{1-\frac{5\rho}{\mu+\nu}}{2-2c_Q\left\|(a^2+2a+2)\alpha\right\|^2} \tag{2.42}$$

hold, we obtain

$$S_2 \ll_q (\mu+\nu)^{\omega(q)+2} Q^{\mu+\nu-\rho}.$$

Together with (2.35) this yields

$$S \ll_\varepsilon Q^{(\mu+\nu)(1+\varepsilon)-\iota\rho/2},$$

whenever (2.27) and (2.42) are satisfied. Finally, we set $\theta_q(\alpha) = \frac{2c_Q \left\|(a^2+2a+2)\alpha\right\|^2}{32}$, $\delta = \frac{2c_Q \left\|(a^2+2a+2)\alpha\right\|^2}{4(2-2c_Q\|(a^2+2a+2)\alpha\|^2)}$, $\beta_1 = \frac{3\theta_q(\alpha)}{2c_Q\|(a^2+2a+2)\alpha\|^2} + \delta$, $\beta_2 = \frac{1-5\theta_q(\alpha)}{2-2c_Q\|(a^2+2a+2)\alpha\|^2} - \delta$ and

$$\rho = \lfloor \theta_q(\alpha)(\mu+\nu) \rfloor.$$

If M satisfies $x^{\beta_1} \leqslant M \leqslant x^{\beta_2}$, then (see (2.26) for the definition of μ and ν)

$$\beta_1 \frac{\log x}{\log Q} \leqslant \frac{\log M}{\log Q} < \mu \leqslant \frac{\log M}{\log Q} + 1 \leqslant \beta_2 \frac{\log x}{\log Q} + 1,$$

and

$$(1-\beta_2)\frac{\log x}{\log Q} + 1 \leqslant \frac{\log \frac{xQ}{M}}{\log Q} < \nu \leqslant \frac{\log \frac{xQ}{M}}{\log Q} + 1 \leqslant (1-\beta_1)\frac{\log x}{\log Q} + 2.$$

Moreover, we have $xQ \leqslant Q^{\mu+\nu} \leqslant xQ^3$. It is easy to verify that $0 < \beta_1 < 1/10$, $1/2 < \beta_2 < 2/3$ (recall that $c_Q < 1/10$). Furthermore, we have that for sufficiently large x the conditions (2.27) and (2.42) are satisfied and we get

$$S \ll_\varepsilon Q^{(\mu+\nu)(1-\theta_q(\alpha)\iota/2+\varepsilon)}. \tag{2.43}$$

Now we can choose ε appropriately to obtain $S \ll_{q,\alpha} x^{1-\hat{\sigma}''_q(\alpha)}$ with $0 < \hat{\sigma}''_q(\alpha) < \theta_q(\alpha)\iota/2$. Employing Lemma 2.14, this finally proves Proposition 2.13.

2.3.4 Final steps in the proof of Theorem 2.1

Theorem 2.1 is a direct consequence of Vaughan's identity (Lemma 2.10) and the estimates for the type I and type II sums (Propositions 2.12 and 2.13). Indeed, we have for $\sigma_q(\alpha) < \min(\sigma'_q(\alpha), \sigma''_q(\alpha))$,

$$\begin{aligned}\sum_{|n|^2 \leqslant N} \Lambda_i(n)\,\mathrm{e}(\alpha s_q(n)) &= \sum_{k \geqslant 0} \sum_{\frac{N}{Q^{k+1}} \leqslant |n|^2 \leqslant \frac{N}{Q^k}} \Lambda_i(n)\,\mathrm{e}(\alpha s_q(n)) \\ &\ll_{q,\alpha} \sum_{k \geqslant 0} \left(\frac{N}{Q^k}\right)^{1-\sigma_q(\alpha)}.\end{aligned}$$

Estimating the geometric series yields the desired result.

2.4 Proofs of Corollary 2.2 and Corollary 2.3

Since the two corollaries follow in a straightforward manner, we only provide sketches of the proofs (compare with [DRS08, Theorems 2.2 and 2.3]). Before we give them, we show an auxiliary lemma concerning congruences in the Gaussian integers.

Lemma 2.22. *Let $q = -a \pm i$, $a \in \mathbb{Z}^+$ and $m \in \mathbb{Z}$ satisfying $m \mid a^2 + 2a + 2$. Furthermore let $b \in \mathbb{Z}$, $z = z_1 + iz_2 \in \mathbb{Z}[i]$ and set $d = (m, q - 1)$. Then we have*

$$z_1 \pm (a+1)z_2 \equiv b \bmod m \quad \textit{if and only if} \quad z \equiv b \bmod d.$$

Here the choice of the sign depends on the sign for $q = -a \pm i$.

Proof. First, we prove the claim that

$$z_1 \pm (a+1)z_2 \equiv b \bmod m \quad \text{if and only if} \quad z_2 \mp (a+1)z_1 \equiv \mp b(a+1) \bmod m. \tag{2.44}$$

We have that $(\mp(a+1), m) \mid (a+1, a^2+2a+2) = (a+1, (a+1)^2+1) = 1$. Thus the left-hand side of (2.44) is equivalent to

$$\mp(z_1 \pm (a+1)z_2)(a+1) \equiv \mp b(a+1) \bmod m.$$

Since $m \mid (a^2+2a+2)$, this is equivalent to the right-hand side of (2.44) and the claim is shown. Next we show that

$$\frac{m}{\delta} = d, \tag{2.45}$$

when δ is defined by $\delta = (m, \overline{q-1})$. Indeed, we can write

$$\frac{m}{\delta} = \frac{(m, a^2+2a+2)}{(m, \overline{q-1})} = \frac{(m, (q-1)(\overline{q-1}))}{(m, \overline{q-1})} = \frac{(m, q-1) \cdot (m, \overline{q-1})}{(m, \overline{q-1})}.$$

We obtain the last inequality from the fact that

$$(q-1, \overline{q-1}) = \begin{cases} 1, & \text{if } a \text{ is odd,} \\ 1+i, & \text{otherwise.} \end{cases}$$

This can be easily shown using the identity

$$(q-1, \overline{q-1}) = (-a \pm i - 1, -a \mp i - 1) = (-a \pm i - 1, \mp 2i) = (-a \pm i - 1, 1+i).$$

As a last preparation note that $\overline{q-1} = \mp i(1 \mp i(a+1))$. Now we can prove the stated result. Let us assume that $z \equiv b \bmod d$. Using (2.45), this is equivalent to $z \equiv b \bmod m/\delta$, where δ is defined as above. Since $(m/\delta, (1 \mp i(a+1))/\delta) = 1$ this is equivalent to

$$(z_1 + iz_2)(1 \mp i(a+1)) \equiv b(1 \mp i(a+1)) \bmod m,$$

i.e., it is equivalent to

$$(z_1 \pm (a+1)z_2) + i(z_2 \mp (a+1)z_1) \equiv b + i(\mp b(a+1)) \bmod m.$$

Recall that $d, b \in \mathbb{Z}$. Thus, we get by our first claim that the last statement is equivalent to

$$z_1 \pm (a+1)z_2 \equiv b \bmod m. \qquad \square$$

Proof of Corollary 2.2. Partial summation yields (here we use the convention, that whenever the index of summation is p, we just sum over Gaussian primes p)

$$\left| \sum_{\substack{|p|^2 \leqslant x \\ \gamma_1 \leqslant \mathrm{Arg}(p) < \gamma_2}} \mathrm{e}\left(\alpha s_q(p)\right) \right| \leqslant \frac{4}{\log x} \max_{\sqrt{x} < t \leqslant x} \left| \sum_{\substack{|p|^2 \leqslant t \\ \gamma_1 \leqslant \mathrm{Arg}(p) < \gamma_2}} \mathrm{e}\left(\alpha s_q(p)\right) \log |p| \right| + O(\sqrt{x}).$$

Using the von Mangoldt function, we obtain

$$\left| \sum_{\substack{|p|^2 \leqslant x \\ \gamma_1 \leqslant \mathrm{Arg}(p) < \gamma_2}} \mathrm{e}\left(\alpha s_q(p)\right) \right| \leqslant \frac{4}{\log x} \max_{t \leqslant x} \left| \sum_{\substack{|n|^2 \leqslant t \\ \gamma_1 \leqslant \mathrm{Arg}(n) < \gamma_2}} \Lambda_i(n)\, \mathrm{e}\left(\alpha s_q(n)\right) \right| + O(\sqrt{x}). \tag{2.46}$$

Let $(p_n)_{n \in \mathbb{N}}$ be the sequence of all Gaussian primes with $\gamma_1 \leqslant \mathrm{Arg}(p_n) < \gamma_2$ ordered such that $|p_{n+1}| \geqslant |p_n|$ (note, that at most eight different primes can have the same absolute value). If α is rational then the sequence $(\alpha s_q(p_n))_{n \in \mathbb{N}}$ is clearly not uniformly distributed. If α is irrational, Theorem 2.1, Hecke's prime number theorem and Weyl's criterion (see Theorem A.18) yield uniform distribution modulo 1. □

Proof of Corollary 2.3. Using Theorem 2.1 (together with (2.46)), one can show exactly the same way as in [DRS08] that

$$\begin{aligned} \#\{|p|^2 \leqslant N :\ p \text{ prime}, \gamma_1 \leqslant \mathrm{Arg}(p) < \gamma_2, s_q(p) \equiv b \bmod g\} \\ = \frac{d}{g} \pi'_{\gamma_1,\gamma_2}(N; b, g) + O_{q,g}(N^{1-\sigma_{q,g}}), \end{aligned}$$

where $\pi'_{\gamma_1,\gamma_2}(N; b, g) = \{|p|^2 \leqslant N : p \text{ prime}, \gamma_1 \leqslant \mathrm{Arg}(p) < \gamma_2, p = z_1 + iz_2, z_1 \pm (a + 1)z_2 \equiv b \bmod d\}$. Here, the choice of the sign depends on the sign of $q = -a \pm i$. It follows from Lemma 2.22 that $\pi'_{\gamma_1,\gamma_2}(N; b, g) = \{|p|^2 \leqslant N : p \text{ prime}, \gamma_1 \leqslant \mathrm{Arg}(p) < \gamma_2, p \equiv b \bmod (d, q-1)\}$. But since

$$(d, q-1) = (g, a^2 + 2a + 2, q - 1) = (g, (q-1)(\overline{q-1}), q-1) = (g, q-1),$$

we are done. □

Chapter 3

The sum of digits of squares in the Gaussian integers

The essence of mathematics lies in its freedom.

Georg Cantor (1845–1918)

In this chapter, we consider the complex sum-of-digits function s_q for squares with respect to special bases q of a canonical number system in the Gaussian integers $\mathbb{Z}[i]$. In particular, we show that the sequence $(\alpha s_q(z^2))_{z\in\mathbb{Z}[i]}$ is uniformly distributed modulo 1 if and only if α is irrational. Furthermore we introduce special sets of Gaussian integers (related to Følner sequences) for which we can determine the order of magnitude of the number of integers z for which $s_q(z^2)$ lies in a fixed residue class $\bmod m$. This extends a recent result of Mauduit and Rivat to $\mathbb{Z}[i]$. The Fourier theoretic results that we obtained in Chapter 2 play again an important role.

3.1 Introduction and main results

The main motivation of the work presented in this chapter is to extend a recent result of Mauduit and Rivat [MR09] on the usual sum-of-digits function s_q, where q denotes an integer $\geqslant 2$. As already discussed in Chapter 1, they answered an open question posed by Gelfond [Gel68] concerning the distribution of $s_q(n^2)$ in arithmetic progressions. Furthermore, they studied the sequence $\left(\alpha s_q(n^2)\right)_{n\in\mathbb{N}}$ and they showed that it is uniformly distributed modulo 1 if and only if $\alpha \in \mathbb{R}\setminus\mathbb{Q}$. One can now ask whether or not these (and similar) results are valid in the case of Gaussian integers.

The specific notation in this chapter is as in the previous one, that is, we write $\mathrm{e}(x)$ for the exponential function $e^{2\pi i x}$, if $q = -a \pm i$ (choose a sign) for a positive integer a, then $Q = |q|^2 = a^2 + 1$, we denote by $R_m\,(\subseteq \mathbb{Z}[i])$ a complete residue system modulo m and write $\|x\|$ for the distance from $x \in \mathbb{R}$ to its nearest integer. Additionally, we denote by $A \triangle B$ the symmetric difference of two sets A and B.

The symbol $f \ll g$ again means that $f = O(g)$. However, this time, if not otherwise stated, the implied constant in the big O-term may depend on q. Recall that if $q = -a \pm i$ and $a \in \mathbb{Z}^+$, then every $z \in \mathbb{Z}[i]$ has a unique finite representation of the form

$$z = \sum_{j \geqslant 0} \varepsilon_j(z) q^j, \qquad \varepsilon_j(z) \in \{0, 1, \ldots, |q|^2 - 1\},$$

with $\varepsilon_j(z) = 0$ for j greater than or equal to some constant $j_0(z)$. In what follows, we denote again by s_q the complex sum-of-digits function (see Appendix A.1).

In Chapter 2 we considered the sum of digits of Gaussian primes. In particular, we showed that the sequence $(\alpha s_q(p))$ running over all Gaussian primes lying in a circular sector is uniformly distributed modulo 1 if and only if α is irrational and we studied the distributional behavior of the sum-of-digits function of primes in arithmetic progressions.

Concerning squares in $\mathbb{Z}[i]$, Gittenberger and Thuswaldner [GT00] dealt with the asymptotic normality of the sum-of-digits function. In particular, they showed that

$$\frac{1}{\#\left\{z : |z|^2 < N\right\}} \# \left\{ |z|^2 < N : \frac{s_q(z^2) - \mu_Q \log_Q N^2}{\sqrt{\sigma_Q^2 \log_Q N^2}} < y \right\} \to \Phi(y)$$

as N goes to infinity, where Φ is the normal distribution function, z runs through the Gaussian integers, $\mu_Q = (Q-1)/2$ and $\sigma_Q^2 = (Q^2-1)/12$. For an precise analysis of this statement see Chapter 7, where we show a local distribution result. Local results of the sum-of-digits function in $\mathbb{Z}[i]$ were treated by Drmota, Grabner and Liardet [DGL08] (again, see Chapter 7 for further information on this subject). In their paper they also used an approach based on ergodic $\mathbb{Z}[i]$-actions and skew products to extend distributional results with respect to so-called Følner sequences $(B_n)_{n \in \mathbb{N}}$. The crucial property of such a sequence is that for all $g \in \mathbb{Z}[i]$ one has $\#(B_n \triangle (g + B_n)) = o(\#B_n)$ (for a complete definition see [DGL08, Section 5]). They showed for example that

$$\lim_{n \to \infty} \frac{1}{\#B_n} \#\{z \in B_n : s_q(z + y) \equiv m \bmod M\} = \frac{1}{M},$$

with respect to any Følner sequence $(B_n)_{n \in \mathbb{N}}$. In what follows, we define κ-$\mathbb{Z}[i]$ sequences which are special Følner sequences where the rate of convergence comes into play.

Definition 3.1. Let κ be a real number satisfying $0 < \kappa \leqslant 1/2$. A sequence $(\mathcal{D}_N)_{N \in \mathbb{N}}$ of subsets of $\mathbb{Z}[i]$ is called a κ-$\mathbb{Z}[i]$ sequence, if for all $N \in \mathbb{N}$,

(i) $\mathcal{D}_N \subseteq \mathcal{D}_{N+1}$,

(ii) $\mathcal{D}_N \subseteq \{z \in \mathbb{Z}[i] : \max(|\Re(z)|, |\Im(z)|) \leqslant \sqrt{N}\}$,

(iii) there exists a constant $c > 0$, such that $cN \leqslant \#\mathcal{D}_N$, and

(iv) $\#\{\mathcal{D}_N \triangle (r + \mathcal{D}_N)\} \ll |r| N^{1-\kappa}$ for all $r \in \mathbb{Z}[i]$.

The simplest examples of $1/2$-$\mathbb{Z}[i]$ sequences are Gaussian integers lying in squares with side length $\sqrt{N}$ or discs with radius $\sqrt{N}$. It turns out that every sequence of convex sets satisfying conditions (i), (ii) and (iii) is a $1/2$-$\mathbb{Z}[i]$ sequence (see Section 3.2).

In this work we show distributional results of the sum-of-digits function of squares with respect to κ-$\mathbb{Z}[i]$ sequences. Our main objective is to obtain information on the exponential sum $\sum_{z\in\mathcal{D}_N} \mathrm{e}(\alpha s_q(z^2))$.

In what follows we have to assume that $q=-a\pm i$ has not too small prime divisors, that is, every divisor $p \mid q$ has to satisfy $|p| \geqslant \sqrt{689}$. This restriction comes into play when considering the Fourier transform of the sum-of-digits function. In contrast to the real case, there arise some technical problems which makes it impossible to cover general q. In particular, it is not possible to use two van der Corput-type inequalities in the same way as Mauduit and Rivat did in the case of squares in $\mathbb{N}$. See the introduction of Section 3.4.3 for a more precise analysis of this problem. Nevertheless, we conjecture that the theorem stated below holds true for all possible bases similar to the results of Chapter 2. We set

$$\begin{aligned}\mathcal{A} :&= \{a \in \mathbb{N} : \text{ if } p \mid q = -a \pm i, \text{ then } |p| \geqslant \sqrt{689}\}\\ &= \{36, 40, 54, 56, 66, 74, 84, 90, 94, \ldots\}.\end{aligned}$$

From the proof of a result of Iwaniec [Iwa78] concerning the representation of a^2+1 as almost-primes, it turns out that this set is infinite.

Theorem 3.2. *Let $a \in \mathcal{A}$ and $\alpha \in \mathbb{R}$. Furthermore let κ be a real number satisfying $0 < \kappa \leqslant 1/2$ and $(\mathcal{D}_N)_{N\in\mathbb{N}}$ a κ-$\mathbb{Z}[i]$ sequence. Then there exists a constant $c_{q,\kappa} > 0$, such that*

$$\sum_{z\in\mathcal{D}_N} \mathrm{e}\left(\alpha s_q(z^2)\right) \ll_q (\log N)^{\omega(q)/2+1} N^{1-c_{q,\kappa}\|(a^2+2a+2)\alpha\|^2},$$

where $\omega(q)$ denotes the number of distinct prime divisors of q.

We can derive from this theorem the order of magnitude of the number of squares whose sum-of-digits function evaluation lies in a fixed residue class modulo some integer m. Furthermore we get that the sequence $\alpha s_q(z^2)$ is uniformly distributed modulo 1 (if and only if $\alpha \in \mathbb{R}\setminus\mathbb{Q}$). We set $Q(b,s) = \#\{z \in R_s : z^2 \equiv b \bmod s\}$.

Corollary 3.3. *Let $a \in \mathcal{A}$ and $b, g \in \mathbb{Z}$, $g \geqslant 2$. Set $d = (g, q-1)$. Furthermore let $(\mathcal{D}_N)_{N\in\mathbb{N}}$ be a κ-$\mathbb{Z}[i]$ sequence (with $0 < \kappa \leqslant 1/2$). Then there exists a constant $\sigma_{q,g,\kappa} > 0$ such that*

$$\#\left\{z \in \mathcal{D}_N : s_q(z^2) \equiv b \bmod g\right\} = \frac{\#\mathcal{D}_N}{g} Q(b,d) + O_{q,g}\left(N^{1-\sigma_{q,g,\kappa}}\right). \tag{3.1}$$

Corollary 3.4. *Let $a \in \mathcal{A}$ and $(z_n)_{n\in\mathbb{N}}$ be a sequence of all Gaussian integers ordered such that $|z_{n+1}| \geqslant |z_n|$. Then the sequence $(\alpha s_q(z_n^2))_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 if and only if α is irrational.*

The chapter is organized as follows. In the next section we discuss properties of κ-$\mathbb{Z}[i]$ sequences defined by convex sets. In Section 3.3 we give a complete treatment of Gauss sums in $\mathbb{Z}[i]$. We prove the main result of this work in Section 3.4, i.e., we derive an estimate of the exponential sum

$$\sum_{z\in\mathcal{D}_N} \mathrm{e}\left(\alpha s_q(z^2)\right). \tag{3.2}$$

The underlying method which we use is based on the work of Mauduit and Rivat [MR09] on the real sum-of-digits function of squares. In a first step (Section 3.4.1) we transform the problem in such a way that we are able to work with differences of the form

$$\alpha s_q((z+r)^2) - \alpha s_q(z^2). \tag{3.3}$$

In order to do so, we prove a van der Corput-type inequality which is the key lemma for working with κ-$\mathbb{Z}[i]$ sequences. It turns out that this lemma is also of great importance for local distribution results proved in Chapter 7.

The advantage of expressions of the form (3.3) is that the higher placed digits of z do not contribute to the exact value in the most cases. With the help of the addition automaton, we show that we can replace the sum-of-digits function with a truncated version of it ("carry lemma"). In Section 3.4.2 we evaluate sums containing linear exponential sums. These results enable us to find an upper bound of (3.2) which only depends on the discrete Fourier transform of the truncated sum-of-digits function. Section 3.4.3 finally contains the last steps of the proof of Theorem 3.2.

In Section 3.5 we first prove a basic property of κ-$\mathbb{Z}[i]$ sequences. With help of this result and Theorem 3.2 we give (for the sake of completeness) straightforward proofs of Corollary 3.3 and Corollary 3.4.

3.2 Convex sets and κ-$\mathbb{Z}[i]$ sequences

In this section we shed some light on κ-$\mathbb{Z}[i]$ sequences.[1] Let us recall that a κ-$\mathbb{Z}[i]$ sequence $(\mathcal{D}_N)_{N\in\mathbb{N}}$ (with $0 < \kappa \leqslant 1/2$) satisfies for all $N \in \mathbb{N}$ the following four conditions:

(i) $\mathcal{D}_N \subseteq \mathcal{D}_{N+1}$,

(ii) $\mathcal{D}_N \subseteq \{z \in \mathbb{Z}[i] : \max\left(|\Re(z)|, |\Im(z)|\right) \leqslant \sqrt{N}\}$,

(iii) there exists a constant $c > 0$, such that $cN \leqslant \#\mathcal{D}_N$, and

(iv) $\#\{\mathcal{D}_N \triangle (r + \mathcal{D}_N)\} \ll |r| N^{1-\kappa}$ for all $r \in \mathbb{Z}[i]$.

[1] The author wants to thank Christoph Haberl and Christian Steineder for ideas presented in this section.

The implied constant in condition (iv) may depend on the sequence but not on r and N. If $|r| \geqslant N^\kappa$, then we have

$$\#\{\mathcal{D}_N \triangle (r+\mathcal{D}_N)\} \leqslant \#\{\mathcal{D}_N\} + \#\{r+\mathcal{D}_N\} \ll N \leqslant |r|N^{1-\kappa}.$$

Thus, one can additionally assume in condition (iv) that $|r| < N^\kappa$. In the definition of a κ-$\mathbb{Z}[i]$ sequence we asked κ to be less or equal to $1/2$. The reason for this demand is the fact that there do not exist κ-$\mathbb{Z}[i]$ sequences with $\kappa > 1/2$. In order to see this, take $r = 1$. If $\#\{\mathcal{D}_N \triangle (1+\mathcal{D}_N)\} \ll N^{1-\kappa}$, then the number of Gaussian integers lying in $\mathcal{D}_N$ is trivially bounded above by $N^{1/2} \cdot N^{1-\kappa} = N^{1-(\kappa-1/2)}$. This contradicts condition (iii). It is clear that every κ-$\mathbb{Z}[i]$ sequence $(\mathcal{D}_N)_{N\in\mathbb{N}}$ is also a κ'-$\mathbb{Z}[i]$ sequence if $\kappa' < \kappa$. Although the value $\kappa = 1/2$ yields the most restrictive case, there exist a big family of $1/2$-$\mathbb{Z}[i]$ sequences. In what follows, we show that every sequence of convex sets satisfying conditions (i), (ii) and (iii) is a $1/2$-$\mathbb{Z}[i]$ sequence. More precisely, let $(C_N)_{N\in\mathbb{N}}$ be a sequence of convex subsets of $\mathbb{C}$ with $C_N \subseteq C_{N+1}$, $C_N \subseteq \{z \in \mathbb{C} : \max(|\Re(z)|, |\Im(z)|) \leqslant \sqrt{N}\}$ and such that the volume of C_N is greater than cN for all $N \in \mathbb{N}$ and for a positive constant $c > 0$. (If we speak about the volume, we mean the 2-dimensional Lebesgue measure of C_N considered as a subset of $\mathbb{R}^2$.) Let $(\mathcal{C}_N)_{N\in\mathbb{N}}$ be defined by

$$\mathcal{C}_N = C_N \cap \mathbb{Z}[i].$$

Proposition 3.5. *The sequence $(\mathcal{C}_N)_{n\in\mathbb{N}}$ is a κ-$\mathbb{Z}[i]$ sequence with $\kappa = 1/2$.*

Conditions (i) and (ii) are trivially satisfied. In order to see that Condition (iii) holds, one only has to use the fact that the number of points lying in $\mathcal{C}_N$ is equal to the volume of C_N (which we denote by $V(C_N)$) plus an error term coming from Gaussian integers lying near the boundary of C_N. This error term can be bounded above by some constant times the Minkowski surface area $S(C_N)$ (see [Gru07, Section 6.4] for the notion of the Minkowski surface area). Since this surface area is monotone, condition (ii) implies that it is bounded by $8N^{1/2}$ and condition (iii) holds true for $(\mathcal{C}_N)_{N\in\mathbb{N}}$. Moreover, the same argument shows that we can assume that the sets C_N are compact and that it suffices to verify that

$$VC_N\triangle(r+C_N)) \ll |r|N^{1/2}$$

in order to confirm condition (iv). We give two different proofs of Proposition 3.5. The first one uses geometric considerations and the second one Steiner's formula for parallel bodies.

First Proof of Proposition 3.5. Without loss of generality, we can assume that $r = ir'$ with $r' \in \mathbb{N}$. We can translate the convex body C_N in such a way that

$$C_N \subseteq \{z \in \mathbb{C} : 0 \leqslant \Re(z), \Im(z) \leqslant 2\sqrt{N}\}.$$

Then there exists two real numbers a and b and two function $f_1(x) \leqslant f_2(x)$, such that the points

$$x + if_1(x) \quad \text{and} \quad x + if_2(x)$$

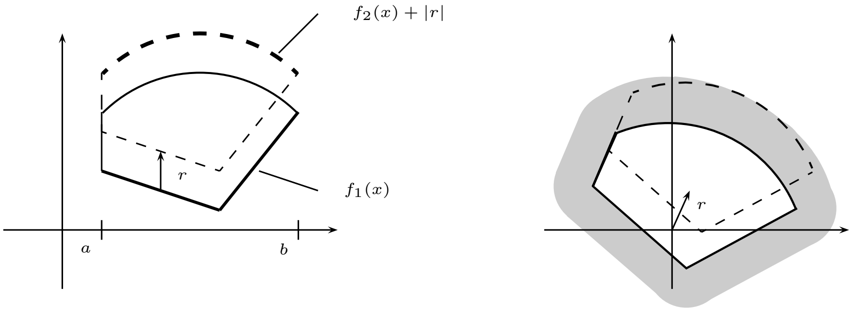


Figure 3.1: The convex sets C_N and $C_N + r$

for $x \in [a, b]$ lie on the boundary of C_N and $C_N \subseteq \{z \in \mathbb{C} : a \leqslant \Re(z) \leqslant b\}$ (see the left-hand side of Figure 3.1). The two functions are almost everywhere differentiable (see [Gru07, Theorem 5.4]) and the volume of $C_N \triangle (r + C_N)$ is bounded above by

$$\left(\int_a^b (f_1(x) + |r|)\mathrm{d}x - \int f_1(x)\mathrm{d}x\right) + \left(\int_a^b (f_2(x) + |r|)\mathrm{d}x - \int f_2(x)\mathrm{d}x\right) = 2\int_a^b |r|\,\mathrm{d}x.$$

Thus, we obtain

$$V(C_N \triangle (r + C_N)) \leqslant 2\int_a^b |r|\,\mathrm{d}x = 2|r|(b-a) \leqslant 4|r|N^{1/2}.$$

As remarked above, this proves the desired result. □

Second Proof of Proposition 3.5. We can assume that the sets C_N are contained in $\mathbb{R}^2$ and that r is a vector in $\mathbb{R}^2$. If C and D are two convex bodies, we denote by $C + D$ the Minkowski sum $C + D = \{u + v : u \in C, v \in D\}$. Let B^2 be the closed unit ball $B^2 = \{x \in \mathbb{R}^2 : \|x\|_2 \leqslant 1\}$, where $\|\cdot\|_2$ denotes the Euclidean norm. Then we have

$$C_N \triangle (r + C_N) \subseteq \Big((C_N + |r|B^2) \setminus C_N\Big) \cup \Big(((C_N + r) + |r|B^2) \setminus (C_N + r)\Big).$$

The gray region on the right-hand side of Figure 3.1 denotes the set $(C_N + |r|B^2) \setminus C_N$. We get

$$\begin{aligned} V(C_N \triangle (r + C_N)) &\leqslant V((C_N + |r|B^2) \setminus C_N) + V(((C_N + r) + |r|B^2) \setminus (C_N + r)) \\ &= 2V(C_N + |r|B^2) - 2V(C_N). \end{aligned}$$

Next, we use Steiner's formula for parallel bodies (see [Gru07, Theorem 6.6]) together with the notion of Minkowski's surface area $S(C)$. We get

$$V(C_N + |r|B^2) = V(C) + S(C)|r| + V(B^2)|r|^2.$$

As we can assume $|r| \leqslant N^{1/2}$, this proves the desired result. □

3.3 Gauss sums in $\mathbb{Z}[i]$

In this section we treat Gauss sums in $\mathbb{Z}[i]$ of the form

$$G(b,\ell;m) = \sum_{n\in R_m} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{bn^2+\ell n}{m}\right)\right),$$

where b, ℓ, $m \in \mathbb{Z}[i]$ with $m \neq 0$ and R_m is a complete residue system modulo m ($\operatorname{tr}(z) = z + \bar{z} = 2\Re(z)$ denotes again the trace of $z \in \mathbb{C}$). Appendix A.3 contains information on classical Gauss sums. The proof of the main result in this section is based on a thorough analysis of the treatment of Gauss sums in the integers given in the book of Graham and Kolesnik [GK91, Chapter 7.4].

Proposition 3.6. *Let b, ℓ, $m \in \mathbb{Z}[i]$ satisfying $m \neq 0$ and $(b,m) = 1$. Then we have*

$$|G(b,\ell;m)| \leqslant 2|m|.$$

This result implies the following (equivalent) result.

Proposition 3.7. *Let b, ℓ, $m \in \mathbb{Z}[i]$ satisfying $m \neq 0$ and set $d = (b,m)$. Then we have*

$$|G(b,\ell;m)| = 0 \qquad \text{if } d \nmid \ell,$$

and

$$|G(b,\ell;m)| \leqslant 2|dm| \qquad \text{if } d \mid \ell.$$

Proof of Proposition 3.7. Set $\tilde{m} = m/d$, $\tilde{b} = b/d$ and $\tilde{\ell} = \ell/d$. Then $\{k\tilde{m} + r : k \in R_d,\ r \in R_{\tilde{m}}\}$ forms a complete residue system modulo m. Thus we can write

$$\begin{aligned}\sum_{n\in R_m} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{bn^2+\ell n}{m}\right)\right) &= \sum_{k\in R_d}\sum_{r\in R_{\tilde{m}}} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{\tilde{b}d(k\tilde{m}+r)^2+\ell(k\tilde{m}+r)}{d\tilde{m}}\right)\right)\\ &= \sum_{r\in R_{\tilde{m}}} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{\tilde{b}r^2+\tilde{\ell}r}{\tilde{m}}\right)\right)\sum_{k\in R_d} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{\ell k}{d}\right)\right).\end{aligned}$$

In the case that $d \nmid \ell$, the inner sum is zero. Otherwise, this sum is equal to $|d|^2$ and we get (using Proposition 3.6)

$$|G(b,\ell;m)| = |d|^2|G(\tilde{b},\tilde{\ell};\tilde{m})| \leqslant 2|d|^2|\tilde{m}| = 2|dm|. \qquad \square$$

In order to prove Proposition 3.6, we show some general properties of Gauss sums. The next two lemmas are straightforward (as in the real case).

Lemma 3.8. *If $(m_1,m_2) = 1$, then we have*

$$G(b,\ell;m_1m_2) = G(b\,m_1,\ell;m_2)G(b\,m_2,\ell;m_1).$$

Proof. Note, that if the numbers j and k run through a complete residue system modulo m_1, resp. m_2, then the numbers $j\,m_2 + k\,m_1$ run through a complete residue system modulo $m_1 m_2$. We obtain

$$\sum_{n\in R_{m_1 m_2}} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{bn^2+\ell n}{m_1 m_2}\right)\right)$$
$$= \sum_{j\in R_{m_1}} \sum_{k\in R_{m_2}} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{b(jm_2+km_1)^2+\ell(jm_2+km_1)}{m_1 m_2}\right)\right),$$

which implies our desired result. □

Lemma 3.9. *Suppose that* $(b, m) = 1$ *and* $(m, -1 + i) = 1$ *or* $(\ell, 2) = 2$*. Then we have*

$$|G(b, \ell; m)| = |G(b, 0; m)|.$$

Proof. First we consider the case $(m, -1 + i) = 1$. Then we have $(4b, m) = 1$ and $4b$ has an inverse element modulo m, say $\tilde{b}$. Replacing n by $n + 2\tilde{b}\ell$ in the index of summation yields

$$G(b, \ell; m) = \sum_{n\in R_m} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{b(n-2\tilde{b}\ell)^2+\ell(n-2\tilde{b}\ell)}{m}\right)\right)$$
$$= \mathrm{e}\left(-\frac{1}{2}\operatorname{tr}\left(\frac{\tilde{b}\ell^2}{m}\right)\right) \sum_{n\in R_m} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{bn^2}{m}\right)\right).$$

In the second case we denote the inverse element of b modulo m by $\hat{b}$. Replacing n by $n + \hat{b}\ell/2$, we can write

$$G(b, \ell; m) = \sum_{n\in R_m} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{b(n-\hat{b}\ell/2)^2+\ell(n-\hat{b}\ell/2)}{m}\right)\right)$$
$$= \mathrm{e}\left(-\frac{1}{2}\operatorname{tr}\left(\frac{\hat{b}\ell^2}{4m}\right)\right) \sum_{n\in R_m} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{bn^2}{m}\right)\right).$$

Thus, we also have $|G(b, \ell; m)| = |G(b, 0; m)|$ in this case. □

In view of Lemma 3.8, it suffices to consider Gauss sums where $m \in \mathbb{Z}[i]$ is a prime power. The set $\{x + iy : 0 \leqslant x < |d + ie|^2/(d, e), 0 \leqslant y < (d, e)\}$ is a complete residue system modulo $d + ie$ (see e.g. [AS03, Theorem 3.10.4]). If $q = a + ib$ is a Gaussian prime with $a \neq 0$, $b \neq 0$ and $|p|^2 > 2$, then we have that $\{n : 0 \leqslant n < |q|^{2r}\}$ already forms a complete residue system modulo q^r. Indeed, if $\delta = (\mathcal{R}(q^r), \mathcal{I}(q^r)) \neq 1$, then $\delta \mid q^r$ and therefore $\delta = q^{r'}$ for some $r' \leqslant r$. But since $q = a + ib$ is a prime with $|q|^2 > 2$, it is impossible that $q^{r'} \in \mathbb{Z}$.

Lemma 3.10. *Suppose that* p *is a prime with* $|p|^2 > 2$*, and* $(b, p) = 1$*. Then we have for* $r \geqslant 1$

$$|G(b, 0; p^r)| = |p|^r.$$

Proof. First we consider the case that p is a Gaussian prime such that $|p|^2$ is a prime number in $\mathbb{Z}$. Then we have

$$\begin{aligned} G(b,0;p^r) &= \sum_{0\leqslant n<|p|^{2r}} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{bn^2}{p^r}\right)\right) = \sum_{0\leqslant n<|p|^{2r}} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{bn^2\bar{p}^r}{|p|^{2r}}\right)\right) \\ &= \sum_{0\leqslant n<|p|^{2r}} \mathrm{e}\left(\frac{cn^2}{|p|^{2r}}\right), \end{aligned}$$

where $c = \Re(\bar{p}^r b)$. Next we show that c and $|p|^2$ are coprime. Set $d = \Im(\bar{p}^r b)$. If $(c,|p|^2) \neq 1$, then $(c,|p|^2) = |p|^2$. This implies that $p \mid c$, $\bar{p} \mid c$ and $\bar{p} \mid d$, which in turn implies that $p \mid d$ and thus $p \mid (c+id)$. It follows that $p \mid \bar{p}$, which is impossible since $|p|^2 > 2$. This allows us to use well-known results for quadratic Gauss sums in $\mathbb{Z}$. Employing [GK91, Lemmas 7.12, 7.13 and 7.15], we obtain

$$|G(b,0;p^r)| = \left|\sum_{0\leqslant n<|p|^{2r}} \mathrm{e}\left(\frac{cn^2}{|p|^{2r}}\right)\right| = \sqrt{|p|^{2r}}.$$

In the second case ($|p|$ is as well a prime in $\mathbb{Z}[i]$ as in $\mathbb{Z}$), the set $\{x+iy : 0 \leqslant x < |p|^r, 0 \leqslant y < |p|^r\}$ forms a complete residue system modulo p^r. First, we consider only primes p, such that $p = |p|$. Put $b = b_1 + ib_2$. If $(b_1,p) = 1$, we have

$$\begin{aligned} \left|\sum_{x=0}^{p^r-1}\sum_{y=0}^{p^r-1} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{b(x+iy)^2}{p^r}\right)\right)\right| &= \left|\sum_{x=0}^{p^r-1}\sum_{y=0}^{p^r-1} \mathrm{e}\left(\frac{b_1(x^2-y^2)-2b_2xy}{p^r}\right)\right| \\ &= \left|\sum_{x=0}^{p^r-1} \mathrm{e}\left(\frac{b_1x^2}{p^r}\right)\sum_{y=0}^{p^r-1} \mathrm{e}\left(\frac{-b_1y^2-2b_2xy}{p^r}\right)\right|. \end{aligned}$$

Using [GK91, Lemma 7.11], we obtain

$$|G(b,0;p^r)| = \left|\sum_{x=0}^{p^r-1} \mathrm{e}\left(\frac{b_1x^2}{p^r}\right)\right| \cdot \left|\sum_{y=0}^{p^r-1} \mathrm{e}\left(\frac{-b_1y^2}{p^r}\right)\right| = p^r.$$

Next we deal with the case $(b_1,p) \neq 1$. We get (setting $p^s = (b_1,p^r)$ and $\widehat{b_1} = b_1p^{-s}$)

$$\begin{aligned} G(b,0;p^r) &= \sum_{x=0}^{p^r-1}\sum_{y=0}^{p^r-1} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{b(x+iy)^2}{p^r}\right)\right) \\ &= \sum_{x=0}^{p^r-1} \mathrm{e}\left(\frac{b_1x^2}{p^r}\right)\sum_{m=0}^{p^{r-s}-1}\sum_{k=0}^{p^s-1} \mathrm{e}\left(\frac{-b_1(kp^{r-s}+m)^2-2b_2x(kp^{r-s}+m)}{p^r}\right) \\ &= \sum_{x=0}^{p^r-1} \mathrm{e}\left(\frac{b_1x^2}{p^r}\right)\sum_{k=0}^{p^s-1} \mathrm{e}\left(\frac{-2b_2xk}{p^s}\right)\sum_{m=0}^{p^{r-s}-1} \mathrm{e}\left(\frac{-\widehat{b_1}m^2-2b_2xmp^{-s}}{p^{r-s}}\right). \end{aligned}$$

If $p^s \nmid x$, then the sum over k equals zero. If $p^s \mid x$, then it is equal to p^s. Indeed, we have $(p,2)=1$ and $(p,b_2)=1$ (note, that otherwise $(b,p)\neq 1$) which implies that $(2b_2,x)=1$. We obtain

$$G(b,0;p^r) = p^s \sum_{x=0}^{p^{r-s}-1} \mathrm{e}\left(\frac{\widehat{b_1}p^{2s}x^2}{p^{r-s}}\right) \sum_{m=0}^{p^{r-s}-1} \mathrm{e}\left(\frac{-\widehat{b_1}m^2 - 2b_2xm}{p^{r-s}}\right).$$

Since p is an odd prime satisfying $p \equiv 3 \bmod 4$ we have for $\lambda \geqslant 1$

$$\sum_{n=0}^{p^\lambda-1} \mathrm{e}\left(\frac{an^2+\ell n}{p^\lambda}\right) = \begin{cases} \mathrm{e}\left(-\frac{\tilde{a}\ell^2}{p^\lambda}\right) p^{\lambda/2}, & \text{if } \lambda \text{ is even,} \\ i\left(\frac{a}{p}\right) \mathrm{e}\left(-\frac{\tilde{a}\ell^2}{p^\lambda}\right) p^{\lambda/2}, & \text{otherwise,} \end{cases}$$

if $(a,p)=1$. Here, $\left(\frac{a}{p}\right)$ denotes the Legendre symbol and $\tilde{a}$ is the inverse element of $4a$ modulo p^λ. This fact follows from [GK91, Lemmas 7.11, 7.12, 7.13 and 7.15]. Thus we get

$$\begin{aligned} |G(b,0;p^r)| &= p^{(r+s)/2} \left| \sum_{x=0}^{p^{r-s}-1} \mathrm{e}\left(\frac{(\widehat{b_1}p^{2s} - 4\tilde{b}b_2^2)x^2}{p^{r-s}}\right) \right| \\ &= p^{(r+s)/2} p^{(r-s)/2} = p^r. \end{aligned}$$

The number $\tilde{b}$ denotes the inverse of $-4\widehat{b_1}$ modulo p^{r-s}. Since $(\tilde{b},p)=1$ and $(b_2,p)=1$, we have $(\widehat{b_1}p^{2s} - 4\tilde{b}b_2^2, p) = 1$ which verifies the last calculation. Since the obtained result holds for all $b \in \mathbb{Z}[i]$ satisfying $(b,p)=1$, we have shown that $|G(b,o,p^r)| = |p|^r$ for all Gaussian primes p with $p=|p|$. □

Next we consider the case $|p|^2=2$. In what follows we fix $p=-1+i$. Recall, that $\mathcal{F}_\lambda = \{\sum_{j=0}^{\lambda-1} \varepsilon_j(-1+i)^j : \varepsilon_j \in \{0,1\}\}$ forms a complete residue system modulo p^λ. In the case where we consider small powers of p, one can easily check following results: Setting $b=b_1+ib_2$ and $\ell=\ell_1+i\ell_2$, we have for $(b,p)=1$ (this is equivalent to $b_1 \not\equiv b_2 \bmod 2$)

$$|G(b,\ell;p)| = \begin{cases} 2, & \text{if } \ell_1 \not\equiv \ell_2 \bmod 2, \\ 0, & \text{if } \ell_1 \equiv \ell_2 \bmod 2, \end{cases}$$

$$|G(b,\ell;p^2)| = \begin{cases} 0, & \text{if } \ell_1 \not\equiv \ell_2 \bmod 2, \\ 0, & \text{if } \ell_1 \equiv \ell_2 \equiv 1 \bmod 2 \text{ and } b_1 \equiv 1 \bmod 2 \text{ or} \\ & \quad \ell_1 \equiv \ell_2 \equiv 0 \bmod 2 \text{ and } b_1 \equiv 0 \bmod 2, \\ 4, & \text{if } \ell_1 \equiv \ell_2 \equiv 1 \bmod 2 \text{ and } b_1 \equiv 0 \bmod 2 \text{ or} \\ & \quad \ell_1 \equiv \ell_2 \equiv 0 \bmod 2 \text{ and } b_1 \equiv 1 \bmod 2, \end{cases}$$

and

$$|G(b,\ell;p^3)| = \begin{cases} 0, & \text{if } \ell_1 \not\equiv \ell_2 \bmod 2, \\ \sqrt{2}\cdot 4, & \text{if } \ell_1 \equiv \ell_2 \equiv 1 \bmod 2, \\ 0, & \text{if } \ell_1 \equiv \ell_2 \equiv 0 \bmod 2. \end{cases}$$

Thus we have $|G(b, \ell; p^r| \leqslant 2|p|^r$ for $r = 1, 2, 3$ and all b, $\ell \in \mathbb{Z}[i]$, $(b, p) = 1$. In general, we get the same kind of estimate. The following lemma implies $|G(b, \ell; p^r)| \leqslant 2|p|^r$ for $r \geqslant 1$ and b, $\ell \in \mathbb{Z}[i]$ with $(b, p) = 1$:

Lemma 3.11. *Let $p \in \mathbb{Z}[i]$ be prime, with $|p|^2 = 2$ and b, $\ell \in \mathbb{Z}[i]$ satisfying $(b, p) = 1$. Then we have for $r \geqslant 4$*

$$|G(b, \ell; p^r)| = \begin{cases} 0, & \textit{if } \ell_1 \textit{ or } \ell_2 \textit{ is odd,} \\ 2 \cdot 2^{r/2}, & \textit{otherwise.} \end{cases}$$

Proof. First we show the result for $p = -1 + i$. We can write

$$\begin{aligned} G(b, \ell; p^r) &= \sum_{n \in \mathcal{F}_r} \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{bn^2 + \ell n}{p^r}\right)\right) \\ & \sum_{j \in \mathcal{F}_1} \sum_{k \in \mathcal{F}_{r-1}} \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{b(jp^{r-1} + k)^2 + \ell(jp^{r-1} + k)}{p^r}\right)\right) \\ & \sum_{k \in \mathcal{F}_{r-1}} \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{bk^2 + \ell k}{p^r}\right)\right) \sum_{j=0}^{1} \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{\ell \bar{p} j}{2}\right)\right). \end{aligned}$$

Since the inner sum is 0 if ℓ_1 and ℓ_2 are from different parity, we have the desired result in this case. If ℓ_1 and ℓ_2 have the same parity, then the internal sum is 2 and we derive

$$G(b, \ell; p^r) = 2 \sum_{k \in \mathcal{F}_{r-1}} \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{bk^2 + \ell k}{p^r}\right)\right). \tag{3.4}$$

Suppose first, that ℓ_1 and ℓ_2 are both odd. Then we obtain

$$\begin{aligned} G(b, \ell; p^r) &= 2 \sum_{j=0}^{1} \sum_{k \in \mathcal{F}_{r-2}} \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{b(jp^{r-2} + k)^2 + \ell(jp^{r-2} + k)}{p^r}\right)\right) \\ &= 2 \sum_{k \in \mathcal{F}_{r-2}} \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{bk^2 + \ell k}{p^r}\right)\right) \sum_{j=0}^{1} \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{\ell j}{p^2}\right)\right) \end{aligned}$$

and the inner sum is 0. Finally, we focus on the case where ℓ_1 and ℓ_2 are both even. It follows from Lemma 3.9 that we can assume that $\ell_1 = \ell_2 = 0$. Using (3.4), we obtain

$$\begin{aligned} |G(b, \ell; p^r)| &= 2 \left| \sum_{k \in \mathcal{F}_{r-1}} \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{bk^2}{p^r}\right)\right) \right| \\ &= 2 \left| \sum_{k \in \mathcal{F}_{r-2}} \left(\mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{b(pk)^2}{p^r}\right)\right) + \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{b(pk+1)^2}{p^r}\right)\right)\right) \right| \\ &= 2 \left| G(b, 0, p^{r-2}) + \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{b}{p^r}\right)\right) G(b, b\bar{p}, p^{r-2}) \right|. \end{aligned}$$

If $r = 4$ we have to use the already calculated values for $G(b, \ell; p^2)$, where the parity of b_1 is important. However, in both cases (b_1 equal or odd) one of the sums above is zero and the other one has absolute value 4. If $r = 5$, we use the values of $G(b, \ell; p^3)$ and for $r \geqslant 6$ a simple induction argument works. In any case, we obtain that

$$|G(b, \ell; (-1+i)^r)| = \begin{cases} 0, & \text{if } \ell_1 \text{ or } \ell_2 \text{ is odd,} \\ 2 \cdot 2^{r/2}, & \text{otherwise.} \end{cases}$$

If $p \neq -1+i$ but $|p|^2 = 2$, we use that $\mathcal{F}_r = \{\sum_{j=0}^{r-1} \varepsilon_j (-1+i)^j : \varepsilon_j \in \{0,1\}\}$ is a complete residue system modulo p^r, too. Thus we have

$$\begin{aligned} G(b, \ell; p^r) &= \sum_{n \in \mathcal{F}_r} \mathrm{e}\left(\frac{1}{2} \operatorname{tr} \left(\frac{\varepsilon^r b n^2 + \varepsilon^r \ell n}{(-1+i)^r} \right) \right) \\ &= G(\varepsilon^r b, \varepsilon^r \ell; (-1+i)^r), \end{aligned}$$

where ε is the appropriate unity $\varepsilon \in \{\pm 1, \pm i\}$, such that $\varepsilon p = (-1+i)$. This finally shows the desired result. □

Proof of Proposition 3.6. Employing Lemmas 3.9, 3.10 and 3.11, we finally get for an arbitrary prime number $p \in \mathbb{Z}[i]$ and for $b,\ \ell \in \mathbb{Z}[i]$ with $(b, p) = 1$ the estimate

$$|G(b, \ell; p^r)| \leqslant 2|p|^r.$$

Using this fact combined with Lemma 3.8 finishes the proof of Proposition 3.6. □

3.4 Proof of Theorem 3.2

Recall that we have $q = -a \pm i$ with $a \in \mathbb{Z}^+$ and $\alpha \in \mathbb{R}$. We set $f(n) = \alpha s_q(n)$ and define (as in Chapter 2) for every real positive number B the set $\Xi_B \subset \mathbb{Z}[i]$ by

$$\Xi_B := \{n \in \mathbb{Z}[i] : \max(|\Re(n)|, |\Im(n)|) \leqslant B^{1/2}\}.$$

Let $\nu \in \mathbb{Z}^+$ be defined by $Q^{\nu-1} < N \leqslant Q^\nu$. Then we will show that

$$S := \sum_{n \in \mathcal{D}_N} \mathrm{e}(f(n^2)) \ll_q \nu^{\omega(q)/2+1} Q^{(1 - c_{q,\kappa} ||(a^2+2a+2)\alpha||^2)\nu}, \tag{3.5}$$

where $c_{q,\kappa}$ is a positive constant and $\omega(q)$ denotes the number of distinct prime divisors of q. Of course, this proves Theorem 3.2.

3.4.1 Van der Corput-type inequality and "carry lemma"

We begin with a van der Corput-type inequality for Gaussian integers (different to the one given in Chapter 2). It allows us to work with differences of the form $\alpha s_q((z+r)^2) - \alpha s_q(z^2)$, which is important for the further steps of the proof. Moreover, it has the remarkable property that we can enlarge the domain of summation without producing a too big error term. On the left-hand side of inequality (3.6) we sum up over Gaussian integers lying in $\mathcal{D}_N$, while on the right-hand side we sum up over all Gaussian integers lying in a square which contains $\mathcal{D}_N$. It permits to calculate special exponential sums that appear in later parts of the proof (see Lemma 3.14). The idea of the Lemma and the proof of it is inspired by [MR09, Lemma 15].

Lemma 3.12. *Let B be a real number and N a positive integer satisfying $N \leqslant B$. Furthermore let z_n ($n \in \mathbb{Z}[i]$) be complex numbers with absolute value $\leqslant 1$ and $(\mathcal{D}_M)_{M\in\mathbb{N}}$ be a κ-$\mathbb{Z}[i]$ sequence. Then we have for any real number $R \geqslant 1$,*

$$\left|\sum_{n\in\mathcal{D}_N} z_n\right| \ll \left(\frac{N}{R^2}\sum_{|r|\leqslant 2R}\left(1-\frac{|r|}{2R+1}\right)\left|\sum_{n,n+r\in\Xi_B} z_{n+r}\overline{z_n}\right|\right)^{1/2} + N^{1-\kappa}R. \tag{3.6}$$

Proof. Before we start the main part of the proof, we observe that $\#\mathcal{D}_N \leqslant \#\Xi_N \ll N$. Next we take for convenience $z_n = 0$ for $n \notin \Xi_B$ and put $T(R) = \#\{0 \leqslant |r| \leqslant R\}$. Since the absolute values of the considered complex numbers are $\leqslant 1$, we obtain (using the properties of the κ-$\mathbb{Z}[i]$ sequence)

$$\begin{aligned}\left|T(R)\sum_{n\in\mathcal{D}_N} z_n - \sum_{|r|\leqslant R}\sum_{n\in\mathcal{D}_N} z_{n+r}\right| &\leqslant \sum_{|r|\leqslant R}\left|\sum_{n\in\mathcal{D}_N} z_n - \sum_{n\in\mathcal{D}_N} z_{n+r}\right| \\ &\leqslant \sum_{|r|\leqslant R}\#\{\mathcal{D}_N \triangle (r+\mathcal{D}_N)\} \\ &\ll \sum_{|r|\leqslant R}|r|N^{1-\kappa} \ll T(R)RN^{1-\kappa}.\end{aligned}$$

Thus, we have

$$\left|\sum_{n\in\mathcal{D}_N} z_n\right| \ll \frac{1}{T(R)}\sum_{n\in\mathcal{D}_N}\left|\sum_{|r|\leqslant R} z_{n+r}\right| + N^{1-\kappa}R.$$

Using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}\left(\sum_{n\in\mathcal{D}_N}\left|\sum_{|r|\leqslant R} z_{n+r}\right|\right)^2 &\leqslant (\#\mathcal{D}_N)\cdot\sum_{n\in\mathcal{D}_N}\left|\sum_{|r|\leqslant R} z_{n+r}\right|^2 \\ &\leqslant N\sum_{0\leqslant|r_1|,|r_2|\leqslant R}\;\sum_{n\in\mathbb{Z}[i]} z_{n+r_1}\overline{z_{n+r_2}} \\ &= N\sum_{0\leqslant|r|\leqslant 2R} w(r)\sum_{n\in\mathbb{Z}[i]} z_{n+r}\overline{z_n},\end{aligned}$$

where $w(r) = \#\{(r_1,r_2), 0 \leqslant |r_1|, |r_2| \leqslant R : r_1 - r_2 = r\} \leqslant (2R+1)(2R+1-|r|)$. Since $T(R) \gg R^2$, we finally get

$$\left|\sum_{n\in\mathcal{D}_N} z_n\right| \ll \left(\frac{N}{R^2}\sum_{|r|\leqslant 2R}\left(1-\frac{|r|}{2R+1}\right)\left|\sum_{n,n+r\in\Xi_B} z_{n+r}\overline{z_n}\right|\right)^{1/2} + N^{1-\kappa}R.$$

This shows the desired result. □

We employ Lemma 3.12 with $B = Q^\nu$ and $R = |q|^\rho/2$, where ρ is an integer satisfying

$$2 \leqslant \rho \leqslant \nu/3, \tag{3.7}$$

and get

$$\begin{aligned} S &\ll \left(\frac{N}{Q^\rho} \sum_{|r|\leqslant|q|^\rho} \left(1 - \frac{|r|}{|q|^\rho+1}\right) \left| \sum_{n,n+r\in\Xi_{Q^\nu}} \mathrm{e}(f((n+r)^2) - f(n^2)) \right| \right)^{1/2} + N^{1-\kappa}|q|^\rho \\ &\ll \left(Q^{2\nu-\rho} + Q^\nu \max_{1\leqslant|r|\leqslant|q|^\rho} \left| \sum_{n,n+r\in\Xi_{Q^\nu}} \mathrm{e}(f((n+r)^2) - f(n^2)) \right| \right)^{1/2} + Q^{\nu-(\nu\kappa-\rho/2)}. \end{aligned}$$

In the last step, we separated the case $r = 0$ and $r \neq 0$. Additionally, we get an error term $O(Q^{(3\nu+\rho)/2})$ (inside the square root) when removing the summation condition $n+r \in \Xi_{Q^\nu}$. But since we have assumed $\rho \leqslant \nu/3$, this term can be neglected. Hence, we obtain

$$S \ll Q^{\nu-\rho/2} + Q^{\nu-(\nu\kappa-\rho/2)} + Q^{\nu/2} \max_{1\leqslant|r|\leqslant|q|^\rho} \left| \sum_{n\in\Xi_{Q^\nu}} \mathrm{e}(f((n+r)^2) - f(n^2)) \right|^{1/2}. \tag{3.8}$$

In a next step, we want to use the fact that we are now dealing with expressions of the form $f((n+r)^2) - f(n^2)$ (similar to the reasoning in Chapter 2). If r is small in comparison to n, the higher placed digits of $(n+r)^2$ and n^2 do not differ in “most” of the cases. In order to show this, we use the same truncated sum-of-digits function already defined in (2.30), namely,

$$f_\lambda(z) = \alpha \sum_{j=0}^{\lambda-1} \varepsilon_j(z),$$

where $\varepsilon_j(z)$, $j \geqslant 0$ are the digits of z in the base-q representation. Recall, that this function is periodic with period q^λ. The next lemma gives an upper bound of the number of cases, where it makes a difference if we use the normal or the truncated sum-of-digits function. The reasoning of the first part of its proof is very similar to the reasoning of the proof of Lemma 2.17. Nevertheless, we also give a detailed proof of this part for the sake of understanding and completeness.

Lemma 3.13. *Let $r \in \mathbb{Z}[i]$ with $|r|^2 < Q^\rho$. We denote by $E(r,\nu,\rho)$ the set of Gaussian integers z such that $z \in \Xi_{Q^\nu}$ and*

$$f((z+r)^2) - f(z^2) \neq f_{\nu+2\rho}((z+r)^2) - f_{\nu+2\rho}(z^2). \tag{3.9}$$

Then we have

$$\# E(r,\nu,\rho) \ll Q^{\nu-\gamma\rho},$$

where $0 < \gamma < 1$ is a constant only depending on q.

Proof. By Propostion A.5 we know that there exists a constant c (only depending on q) such that for all $z \in \Xi_{Q^\nu}$ the number z^2 has $\leqslant 2\nu + c$ digits, $2zr + r^2$ has $\leqslant \nu + \rho + c$ digits and if w has ℓ digits, then we have that $|q|^{l-c} \leqslant |w| \leqslant |q|^{\ell+c}$. We can assume that $\rho > 4c$ (the statement is trivial in the converse case).

Similar to Lemma 2.17, every Gaussian integer $z^2 \in \Xi_{Q^\nu}$ can be uniquely written as

$$z^2 = m_0(z^2) + q^{\nu+\rho+c} m_1(z^2) + q^{\nu+2\rho} m_2(z^2),$$

where $m_0(z^2) \in \mathcal{F}_{\nu+\rho+c}$, $m_1(z^2) \in \mathcal{F}_{\rho-c}$ and $m_2(z^2) \in \mathcal{F}_{\nu+c-2\rho}$. We set

$$T = \{t \in \mathcal{F}_{\rho-c} : \ t = m_1(z^2) \text{ for some } z \in E(r,\nu,\rho)\},$$

and

$$N_t = \{z \in \Xi_{Q^\nu} : \ m_1(z^2) = t \text{ and } z \in E(r,\nu,\rho)\}.$$

As in the proof of Lemma 2.17, one can show that $\# T \ll Q^{\gamma'\rho}$, where γ' is a positive constant satisfying $\gamma' < 1$. In what follows we show

$$\# N_t \ll Q^{\nu-\rho}. \tag{3.10}$$

This implies

$$\# E(r,\nu,\rho) = \sum_{t\in T} \# N_t \ll \sum_{t\in T} Q^{\nu-\rho} \ll Q^{\nu-(1-\gamma')\rho},$$

which implies the desired result with $\gamma = 1-\gamma' > 0$.

First, we fix some $t \in \mathcal{F}_{\rho-c}$ and set

$$S_m := \{z : z^2 = m_0 + q^{\nu+\rho+c} t + q^{\nu+2\rho} m, \ m_0 \in \mathcal{F}_{\nu+\rho+c}\}.$$

This allows us to write $N_t \subseteq \bigcup_{m\in\mathcal{F}_{\nu+c-2\rho}} S_m$, and hence

$$\# N_t \leqslant \sum_{m\in\mathcal{F}_{\nu+c-2\rho}} \# S_m. \tag{3.11}$$

If $m = 0$, we have $S_0 \subseteq \{z : |z|^2 \ll |q|^{\nu+2\rho}\}$, and we obtain that $\#S_0 \ll |q|^{\nu+2\rho}$. For $m \neq 0$, we can bound the cardinality of S_m by

$$\#S_m \ll \frac{|q|^{\nu+\rho}}{|t+q^{\rho-c}m|} + \frac{|q|^{\frac{\nu+\rho}{2}}}{\sqrt{|t+q^{\rho-c}m|}} + 1. \tag{3.12}$$

To see this, consider Figure 3.2. The dashed domain in the first drawing contains the set of all possible squares (the solid circle has radius $|q|^{\nu+\rho+2c}$) and the checkered domain contains the set S_m. One possibility to find an upper bound of the cardinality of S_m is to calculate the area of the checkered domain plus an additional area (if the Gaussian integers lie on the "border"). The second drawing shows this region. For readability, we denote the angle by φ and write

$$a = \sqrt{|q|^{\nu+\rho+c}\left(|t+q^{\rho-c}m| - |q|^c\right)} \quad \text{and} \quad b = \sqrt{|q|^{\nu+\rho+c}\left(|t+q^{\rho-c}m| + |q|^c\right)}.$$

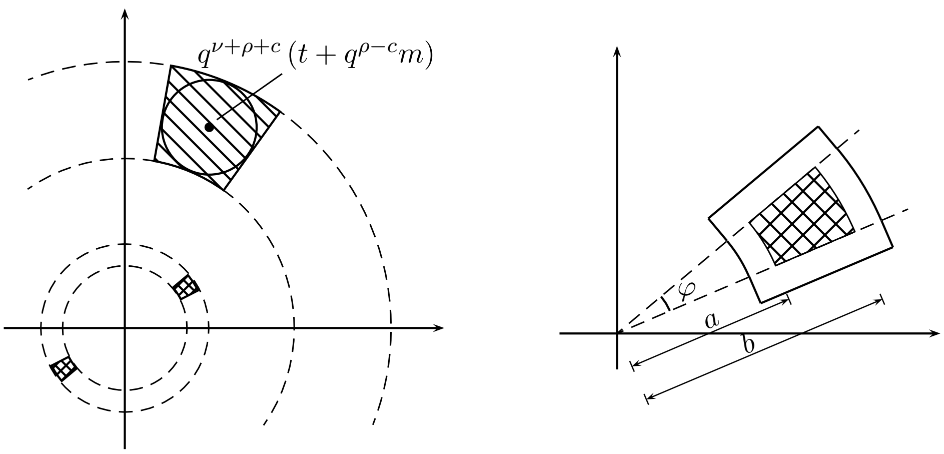


Figure 3.2: Calculation of S_m

First we calculate the area of the circular ring segment. The inner radius is $a-1/\sqrt{2}$ and the outer radius is $b+1/\sqrt{2}$. We get the area

$$(b^2-a^2+\sqrt{2}(b+a))\frac{\varphi}{2} \leqslant (b^2-a^2+2\sqrt{2}b)\varphi. \tag{3.13}$$

The remaining domain (thickness $1/\sqrt{2}$) has area

$$2(b-a+\sqrt{2})\frac{1}{\sqrt{2}} \ll \frac{b^2-a^2}{a}+1. \tag{3.14}$$

Furthermore we have $\varphi \ll |t+q^{\rho-c}m|^{-1}$ (note that $|t+q^{\rho-c}m| \geqslant |q|^{\rho-2c} > |q|^{2c}$). Adding the expressions (3.13) and (3.14) together and using the estimates above, we finally obtain (3.12).

In order to avoid problems arising from the denominators, we split the sum in (3.11) up into two parts. The first one contains all integers m satisfying $|m|^2 \leqslant |q|^{\nu-3\rho}$ (they are all in $\mathcal{F}_{\nu-2\rho+c}$). From (3.12) follows, that $\# S_m \ll |q|^{\nu+\rho}$ for all $m \neq 0$. Using this crude upper bound, we obtain

$$\sum_{0<|m|^2\leqslant|q|^{\nu-3\rho}} S_m \ll \sum_{0<|m|^2\leqslant|q|^{\nu-3\rho}} |q|^{\nu+\rho} \ll |q|^{2\nu-2\rho} = Q^{\nu-\rho}.$$

Now, we evaluate the remaining part in (3.11). Since $|m|^2 > |q|^{\nu-3\rho}$, we have that $|t+q^{\rho-c}m| \gg |q|^{\rho}|m|$. Using (3.12), we get

$$\sum_{\substack{m\in\mathcal{F}_{\nu-2\rho+c}\\ |m|^2>|q|^{\nu-3\rho}}} \# S_m \ll \sum_{\substack{m\in\mathcal{F}_{\nu-2\rho+c}\\ |m|^2>|q|^{\nu-3\rho}}} \left(\frac{|q|^{\nu+\rho}}{|t+q^{\rho-c}m|} + \frac{|q|^{\frac{\nu+\rho}{2}}}{\sqrt{|t+q^{\rho-c}m|}} + 1\right),$$

and hence

$$\sum_{\substack{m\in\mathcal{F}_{\nu-2\rho+c}\\|m|^2>|q|^{\nu-3\rho}}} \# S_m \ll \sum_{\substack{m\in\mathcal{F}_{\nu-2\rho+c}\\|m|^2>|q|^{\nu-3\rho}}} \left(\frac{|q|^\nu}{|m|}+\frac{|q|^{\nu/2}}{|m|^{1/2}}+1\right)$$
$$\ll \sum_{0<|m|^2\leqslant|q|^{2\nu-4\rho+2c}} \left(\frac{|q|^\nu}{|m|}+\frac{|q|^{\nu/2}}{|m|^{1/2}}+1\right).$$

Thus we have to deal with sums of the form $\sum_{0<|m|\leqslant N}\frac{1}{|m|^\alpha}$, where $\alpha\in\{1/2,1\}$. A first crude estimation shows that

$$\sum_{0<|m|\leqslant N}\frac{1}{|m|^\alpha}\ll\sum_{r=1}^{\lfloor N\rfloor} r^{1-\alpha}.$$

Calculating the sum on the right-hand side, we get

$$\sum_{0<|m|\leqslant N}\frac{1}{|m|^\alpha}\ll\begin{cases} N, & \text{if } \alpha=1,\\ N^{3/2}, & \text{if } \alpha=1/2.\end{cases}$$

Hence we obtain

$$\sum_{0<|m|^2\leqslant|q|^{2\nu-4\rho+2c}} \left(\frac{|q|^\nu}{|m|}+\frac{|q|^{\nu/2}}{|m|^{1/2}}+1\right)$$
$$\ll |q|^\nu|q|^{\nu-2\rho}+|q|^{\nu/2}|q|^{\frac{3\nu-6\rho}{2}}+|q|^{2\nu-4\rho}\ll Q^{\nu-\rho}.$$

This shows estimate (3.10) and the proof of Lemma 3.13 is finished. □

For further considerations, we set

$$\lambda=\nu+2\rho. \tag{3.15}$$

Replacing f by f_λ gives a total error of $O(Q^{\nu-\gamma\rho/2})$. Thus, by (3.8) we have

$$S\ll Q^{\nu-\gamma\rho/2}+Q^{\nu-(\nu\kappa-\rho/2)}+Q^{\nu/2}\max_{1\leqslant|r|\leqslant|q|^\rho}|S_1(r,\nu,\rho)|^{1/2}, \tag{3.16}$$

where

$$S_1(r,\nu,\rho)=\sum_{n\in\Xi_{Q^\nu}} \mathrm{e}(f_\lambda((n+r)^2)-f_\lambda(n^2)).$$

3.4.2 Calculations leading to the Fourier transform

Next we estimate some special exponential sums in order to obtain expressions containing Gauss sums and the Fourier transform. In Remark 2.19, we noted that

$$\sum_{z\in\Xi_N}\mathrm{e}\left(\frac{1}{2}\operatorname{tr}((r+is)z)\right)\ll\min\left(N,\frac{\sqrt{N}}{\|r\|},\frac{\sqrt{N}}{\|s\|},\frac{1}{\|r\|\|s\|}\right). \tag{3.17}$$

The following lemma is similar to Lemma 2.20.

Lemma 3.14. *Let $m \in \mathbb{Z}[i] \setminus \{0\}$ and R_m be a complete residue system modulo m. Moreover, let $N \in \mathbb{N}$. Then we have*

$$\sum_{h \in R_m} \min \left(N, \frac{\sqrt{N}}{\|\Re(h/m)\|}, \frac{\sqrt{N}}{\|\Im(h/m)\|}, \frac{1}{\|\Re(h/m)\| \|\Im(h/m)\|} \right) \\ \leqslant 16N + 64\sqrt{N}|m| \log |m| + 64|m|^2 (\log |m|)^2.$$

Remark 3.15. This lemma is an improvement of [GT00, Lemma 2.6], where Gittenberger and Thuswaldner dealt with similar sums. They use the Koksma-Hlawka inequality to obtain an error term of the form $O\left(N|m| + |m|^2 (\log N)^2\right)$, which suffices for the proof of their main result. In our case, the term $N|m|$ is too big and we have to use other ideas in order to succeed.

Proof of Lemma 3.14. Since the summands are periodic with period m, it does not matter over which residue system we sum. Let us denote the considered sum with T. Note that $R = \mathbb{Z}[i] \cap \{(\alpha + i\beta)m : -1/2 \leqslant \alpha, \beta < 1/2\}$ is a complete residue system modulo m and set $\hat{R} = \{(\alpha, \beta) \in [-1/2, 1/2)^2 : (\alpha + i\beta)m \in R\}$. Then, we can write

$$T = \sum_{(\alpha,\beta) \in \hat{R}} \min \left(N, \frac{\sqrt{N}}{\|\alpha\|}, \frac{\sqrt{N}}{\|\beta\|}, \frac{1}{\|\alpha\| \|\beta\|} \right).$$

In a next step, we tessellate the square $[-1/2, 1/2)^2$ with small squares of side length $1/(2 \lfloor |m| \rfloor)$. Therefore, we set

$$\hat{R}_{h,k} := \hat{R} \cap \left[\frac{h}{2 \lfloor |m| \rfloor}, \frac{h+1}{2 \lfloor |m| \rfloor} \right) \times \left[\frac{k}{2 \lfloor |m| \rfloor}, \frac{k+1}{2 \lfloor |m| \rfloor} \right),$$

and

$$\hat{\mathcal{R}}_{h,k} := \hat{R}_{h,k} \cup \hat{R}_{-h-1,k} \cup \hat{R}_{-h-1,-k-1} \cup \hat{R}_{h,-k-1}.$$

We obtain

$$T = \sum_{h=0}^{\lfloor |m| \rfloor - 1} \sum_{k=0}^{\lfloor |m| \rfloor - 1} \sum_{(\alpha,\beta) \in \hat{\mathcal{R}}_{h,k}} \min \left(N, \frac{\sqrt{N}}{\|\alpha\|}, \frac{\sqrt{N}}{\|\beta\|}, \frac{1}{\|\alpha\| \|\beta\|} \right).$$

Note, that the inner sum has less or equal $4(|m|/(2 \lfloor |m| \rfloor) + 1)^2 \leqslant 16$ summands. We distinguish between three different cases. If $h = k = 0$, we use N as an upper bound of the summands. If $h \neq 0$ and $k = 0$, we use $\sqrt{N} \cdot (h/(2 \lfloor |m| \rfloor))^{-1}$ (resp. $\sqrt{N} \cdot (k/(2 \lfloor |m| \rfloor))^{-1}$ when $h = 0$ and $k \neq 0$) and $(h/(2 \lfloor |m| \rfloor))^{-1} \cdot (k/(2 \lfloor |m| \rfloor))^{-1}$ in the remaining case. Thus we have

$$\begin{aligned} T &\leqslant \frac{4|m|^2}{(2 \lfloor |m| \rfloor) + 1)^2} \left(N + 2\sqrt{N} \sum_{k=1}^{\lfloor |m| \rfloor - 1} \frac{1}{k/(2 \lfloor |m| \rfloor)} + \left(\sum_{k=1}^{\lfloor |m| \rfloor - 1} \frac{1}{k/(2 \lfloor |m| \rfloor)} \right)^2 \right) \\ &\leqslant 16N + 64\sqrt{N}|m| \log |m| + 64|m|^2 (\log |m|)^2, \end{aligned}$$

and the desired result follows. □

Lemma 3.16. *Let $q = -a \pm i$ with $a \in \mathbb{Z}^+$, $\lambda \geqslant 0$ and z_n $(n \in \mathbb{Z}[i])$ complex numbers of period q^λ. Then we have for $1 < N \leqslant Q^\lambda$,*

$$\sum_{n \in \Xi_N} z_n \ll \nu^2 \max_{h \in \mathcal{F}_\lambda} \left| \sum_{n \in \mathcal{F}_\lambda} z_n \, \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{hn}{q^\lambda}\right)\right) \right|.$$

Proof. Using the periodicity of the considered complex numbers, we can write

$$\begin{aligned} \sum_{n \in \Xi_N} z_n &= \sum_{n \in \mathcal{F}_\lambda} z_n \sum_{z \in \Xi_N} \frac{1}{Q^\lambda} \sum_{h \in \mathcal{F}_\lambda} \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{h(n-z)}{q^\lambda}\right)\right) \\ &= \frac{1}{Q^\lambda} \sum_{h \in \mathcal{F}_\lambda} \sum_{z \in \Xi_N} \mathrm{e}\left(-\frac{1}{2} \operatorname{tr}\left(\frac{hz}{q^\lambda}\right)\right) \sum_{n \in \mathcal{F}_\lambda} z_n \, \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{hn}{q^\lambda}\right)\right). \end{aligned}$$

Recall that $\lambda = \nu + 2\rho$ and $\rho \leqslant \nu/3$ (see (3.7) and (3.15)). Thus we have $(\log |q|^\lambda)^2 \ll \nu^2$. Applying (3.17) and the previous lemma yields the desired result. □

As indicated above, $f_\lambda(z)$ is periodic with period q^λ. Using the Fourier transform $F_\lambda(h, \alpha)$ (compare with Section 2.2), we obtain

$$S_1 = \sum_{h_1, h_2 \in \mathcal{F}_\lambda} F_\lambda(h_1, \alpha) \overline{F_\lambda(-h_2, \alpha)} \sum_{n \in \Xi_{Q^\nu}} \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{h_1(n+r)^2 + h_2 n^2}{q^\lambda}\right)\right).$$

Employing Lemma 3.16 with $N = Q^\nu$ yields

$$S_1 \ll \nu^2 \sum_{h_1, h_2 \in \mathcal{F}_\lambda} |F_\lambda(h_1, \alpha) F_\lambda(-h_2, \alpha)| \cdot \max_{l \in \mathcal{F}_\lambda} \left| G(h_1 + h_2, 2rh_1 + l; q^\lambda) \right|,$$

where we used the notion of the quadratic Gauss sums (see Section 3.3). Proposition 3.7 gives

$$S_1 \ll \nu^2 Q^{\lambda/2} \max_{l \in \mathcal{F}_\lambda} \sum_{d | q^\lambda} |d| \sum_{\substack{h_1, h_2 \in \mathcal{F}_\lambda \\ (h_1 + h_2, q^\lambda) = d \\ d | 2rh_1 + l}} |F_\lambda(h_1, \alpha) F_\lambda(-h_2, \alpha)|,$$

where we only sum over one of the four possible associated elements of a divisor d.

3.4.3 Final estimates for the proof of Theorem 3.2

For further calculations, we only consider the case where $q = -a \pm i$ with $a \in \mathcal{A}$. This results from the fact that we employ Lemma 2.7 (a Fourier transform estimate), which is not sufficient if q has small prime divisors. In particular, the constant η_k has to be smaller than $1/4$ (several considerations in the following steps of the proof need this assumption). Mauduit and Rivat used two different van der Corput-type inequalities and some further calculations, which allowed them to employ a weaker Fourier transform estimate than the corresponding one we need in our case. Unfortunately, their considerations do not work in our case.

First we replace the summation condition $(h_1 + h_2, k^\lambda) = d$ by the less restrictive one $h_1 + h_2 \equiv 0 \bmod d$. The condition

$$2rh_1 + l \equiv 0 \bmod d \tag{3.18}$$

can be rewritten as follows: Set $\tilde{d} = (2r, d) = (r, d)$ (since $d \mid q^\lambda$, we have $(d, 1+i) = 1$). Then $\tilde{d} \mid l$, since otherwise (3.18) cannot hold. If we set $r' = 2r/\tilde{d}$ and $l' = l/\tilde{d}$, then (3.18) is equivalent to

$$r'h_1 + l' \equiv 0 \bmod d/\tilde{d}.$$

The integer r' has an inverse element modulo $d/\tilde{d}$ which we call r''. If we finally set $l'' = r''l'$, then we can write the last equation as

$$h_1 + l'' \equiv 0 \bmod d/\tilde{d}.$$

We proceed with separating the sum according to the value of $\nu_q(d)$, where $\nu_q(d)$ is the unique integer δ, such that $q^\delta \mid d$ but $q^{\delta+1} \nmid d$. Set

$$\rho_Q := \left\lfloor \frac{\rho \log Q}{\log 689} \right\rfloor. \tag{3.19}$$

Then we can write

$$S_1 \ll \nu^2 Q^{\lambda/2} \max_{l \in \mathcal{F}_\lambda} (S_2 + S_3), \tag{3.20}$$

where

$$S_2 = \sum_{0 \leqslant \delta \leqslant \rho_Q} \sum_{\substack{k \mid q^{\lambda-\delta} \\ q \nmid k}} \left| kq^\delta \right| \sum_{\substack{h_1, h_2 \in \mathcal{F}_\lambda \\ h_1 + h_2 \equiv 0 \bmod kq^\delta \\ h_1 + l'' \equiv 0 \bmod kq^\delta/(kq^\delta, r)}} |F_\lambda(h_1, \alpha) F_\lambda(-h_2, \alpha)|,$$

and

$$S_3 = \sum_{\rho_Q < \delta \leqslant \lambda} \sum_{\substack{k \mid q^{\lambda-\delta} \\ q \nmid k}} \left| kq^\delta \right| \sum_{\substack{h_1, h_2 \in \mathcal{F}_\lambda \\ h_1 + h_2 \equiv 0 \bmod kq^\delta \\ h_1 + l'' \equiv 0 \bmod kq^\delta/(kq^\delta, r)}} |F_\lambda(h_1, \alpha) F_\lambda(-h_2, \alpha)|.$$

In order to find an upper bound of S_2, we replace the condition $h_1 + l'' \equiv 0 \bmod kq^\delta/(kq^\delta, r)$ by the weaker condition

$$h_1 + l'' \equiv 0 \bmod k/(k, r).$$

This is allowed since $k/(k, r)$ is a divisor of

$$\frac{k}{(k, r)} \cdot \frac{(kq^\delta, rq^\delta)}{(kq^\delta, r)} = \frac{kq^\delta}{(kq^\delta, r)}.$$

We can write

$$S_2 \leqslant \sum_{0\leqslant\delta\leqslant\rho_Q} \sum_{\substack{k\mid q^{\lambda-\delta}\\ q\nmid k}} \left|kq^\delta\right| \sum_{\substack{h_1\in\mathcal{F}_\lambda\\ h_1\equiv -l'' \bmod k/(k,r)}} \left|F_\lambda(h_1,\alpha)\right| \sum_{\substack{h_2\in\mathcal{F}_\lambda\\ h_2\equiv -h_1 \bmod kq^\delta}} \left|F_\lambda(-h_2,\alpha)\right|.$$

Since the Fourier transform is periodic with period q^λ in its first argument and since the moduli occurring in the summation conditions are divisors of q^λ, we can sum over an arbitrary complete congruence system modulo q^λ. Using Lemma 2.7 (twice) yields

$$\begin{aligned} S_2 &\leqslant \sum_{0\leqslant\delta\leqslant\rho_Q} \sum_{\substack{k\mid q^{\lambda-\delta}\\ q\nmid k}} \left|kq^\delta\right| Q^{\eta_{689}\lambda} \left(\frac{|k|}{|(k,r)|}\right)^{-2\eta_{689}} Q^{\eta_{689}(\lambda-\delta)} |k|^{-2\eta_{689}} \\ &= Q^{2\eta_{689}\lambda} \sum_{0\leqslant\delta\leqslant\rho_Q} Q^{(1/2-\eta_{689})\delta} \sum_{\substack{k\mid q^{\lambda-\delta}\\ q\nmid k}} |k|^{1-4\eta_{689}} \left|(k,r)\right|^{2\eta_{689}}. \end{aligned}$$

Note, that $|(k,r)| \leqslant |r| \leqslant |q|^\rho$. Furthermore, we have for every $k \mid q^{\lambda-\delta}$ with $q \nmid k$ that

$$|k| = |(k,q^\lambda)| \leqslant |(k,q)|^{\lambda-\delta} \leqslant \left(\frac{|q|}{\sqrt{689}}\right)^{\lambda-\delta} = Q^{\frac{\lambda-\delta}{2}(1-\log_Q 689)}. \tag{3.21}$$

Now we need the first time that $\eta_{689} < 1/4$. If we denote by $\tau(m)$ the number of divisors of m, then we obtain

$$\begin{aligned} S_2 &\leqslant Q^{2\eta_{689}\lambda} \sum_{0\leqslant\delta\leqslant\rho_Q} Q^{\delta(1/2-\eta_{689})} \tau(q^{\lambda-\delta}) Q^{\frac{\lambda-\delta}{2}(1-\log_Q 689)(1-4\eta_{689})} Q^{\rho\eta_{689}} \\ &= \tau(q^\lambda) Q^{\lambda\left(\frac{1}{2}-\frac{1-4\eta_{689}}{2}\log_Q 689\right)+\rho\eta_{689}} \sum_{0\leqslant\delta\leqslant\rho_Q} Q^{\delta\left(\eta_{689}+\frac{1-4\eta_{689}}{2}\log_Q 689\right)}. \end{aligned}$$

Since $\rho \leqslant \rho_Q$ and $(\log_Q 689)(1-4\eta_{689})/2 \leqslant (1-4\eta_{689})/2 < 0.00002$, we get

$$S_2 \ll \tau(q^\lambda) Q^{\lambda\left(\frac{1}{2}-\frac{1-4\eta_{689}}{2}\log_Q 689\right)+\rho_Q}. \tag{3.22}$$

Next we estimate S_3. Since $\delta > \rho_Q$, we have that $(kq^\delta, r) \mid q^{\rho_Q} \mid kq^\delta$. This follows from the fact that $|r| \leqslant |q|^\rho$ and that every prime divisor p of q satisfies

$$\nu_p\left(\left(kq^\delta, r\right)\right) \leqslant \left\lfloor \rho\frac{\log Q}{\log|p|^2} \right\rfloor \leqslant \rho_Q,$$

(cf. [MR09, Section 5.5]). Hence, we are allowed to replace the summation condition $h_1 + l'' \equiv 0 \bmod kq^\delta/(kq^\delta, r)$ by the less restrictive condition $h_1 + l'' \equiv 0 \bmod kq^{\delta-\rho_Q}$. We have

$$S_3 \leqslant \sum_{\rho_Q<\delta\leqslant\lambda} \sum_{\substack{k\mid q^{\lambda-\delta}\\ q\nmid k}} \left|kq^\delta\right| \sum_{\substack{h_1\in\mathcal{F}_\lambda\\ h_1\equiv -l'' \bmod kq^{\delta-\rho_Q}}} \left|F_\lambda(h_1,\alpha)\right| \sum_{\substack{h_2\in\mathcal{F}_\lambda\\ h_2\equiv -h_1 \bmod kq^\delta}} \left|F_\lambda(-h_2,\alpha)\right|.$$

Using Lemma 2.7 in combination with Lemma 2.5, we get

$$\sum_{\substack{h_1\in\mathcal{F}_\lambda\\ h_1\equiv -l'' \bmod kq^{\delta-\rho_Q}}} |F_\lambda(h_1,\alpha)| \sum_{\substack{h_2\in\mathcal{F}_\lambda\\ h_2\equiv -h_1 \bmod kq^{\delta}}} |F_\lambda(-h_2,\alpha)|$$
$$\leqslant |k|^{-4\eta_{689}}\, Q^{(\lambda-\delta+\rho_Q)\eta_{689}-(\delta-\rho_Q)c_Q\left\|(a^2+2a+2)\alpha\right\|^2+(\lambda-\delta)\eta_{689}-\delta c_Q\left\|(a^2+2a+2)\alpha\right\|^2},$$

where c_Q is the positive constant defined in Remark 2.6 (we use this constant rather than the constant defined in Lemma 2.5 since it simplifies further calculations). It follows that

$$S_3 \leqslant Q^{2\eta_{689}\lambda+\rho_Q(\eta_{689}+c_Q\left\|(a^2+2a+2)\alpha\right\|^2)}$$
$$\cdot \sum_{\rho_Q<\delta\leqslant\lambda} Q^{\delta(1/2-2\eta_{689}-2c_Q\left\|(a^2+2a+2)\alpha\right\|^2)} \sum_{\substack{k\mid q^{\lambda-\delta}\\ q\nmid k}} |k|^{1-4\eta_{689}}.$$

By (3.21) and the fact that $\eta_{689} < 1/4$, we have

$$S_3 \leqslant \tau(q^\lambda)Q^{2\eta_{689}\lambda+\rho_Q(\eta_{689}+c_Q\left\|(a^2+2a+2)\alpha\right\|^2)}$$
$$\cdot \sum_{\rho_Q<\delta\leqslant\lambda} Q^{\delta(1/2-2\eta_{689}-2c_Q\left\|(a^2+2a+2)\alpha\right\|^2)+\frac{\lambda-\delta}{2}(1-\log_Q 689)(1-4\eta_{689})}$$
$$= \tau(q^\lambda)Q^{\lambda\left(1/2-\frac{1-4\eta_{689}}{2}\log_Q 689\right)+\rho_Q(\eta_{689}+c_Q\left\|(a^2+2a+2)\alpha\right\|^2)}$$
$$\cdot \sum_{\rho_Q<\delta\leqslant\lambda} Q^{\delta\left(\frac{1-4\eta_{689}}{2}\log_Q 689-2c_Q\left\|(a^2+2a+2)\alpha\right\|^2\right)}.$$

Since

$$2c_Q\left\|(a^2+2a+2)\alpha\right\|^2 \leqslant \frac{1}{2}c_Q \leqslant \frac{\pi^2}{54\cdot 689^2\log Q} < \frac{1-4\eta_{689}}{2}\log_Q 689, \tag{3.23}$$

we obtain

$$S_3 \leqslant \tau(q^\lambda)Q^{\lambda\left(1/2-2c_Q\left\|(a^2+2a+2)\alpha\right\|^2\right)+\rho_Q}. \tag{3.24}$$

Thus we get (see (3.20), (3.22), (3.23) and (3.24)),

$$S_1 \ll \nu^2\tau(q^\lambda)Q^{\lambda\left(1-2c_Q||(a^2+2a+2)\alpha||^2\right)+\rho_Q}.$$

Since $\tau(q^\lambda) \leqslant \lambda^{\omega(q)}\tau(q)$ (where $\omega(q)$ denotes the number of distinct prime divisors of q, cf. (2.41)), we obtain (compare also with (3.15) and (3.19))

$$S_1 \ll \nu^{\omega(q)+2}Q^{\nu\left(1-2c_Q||(a^2+2a+2)\alpha||^2\right)+3\rho\log_{689} Q}.$$

By (3.16) we have

$$S \ll Q^{\nu-\gamma\rho/2} + Q^{\nu-(\nu\kappa-\rho/2)} + Q^{\nu/2}\max_{1\leqslant|r|\leqslant|q|^\rho}|S_1(r,\nu,\rho)|^{1/2}$$
$$\ll Q^{\nu-\gamma\rho/2} + Q^{\nu-(\nu\kappa-\rho/2)} + \nu^{\omega(q)/2+1}Q^{\nu\left(1-c_Q||(a^2+2a+2)\alpha||^2\right)+(3/2)\rho\log_{689} Q}.$$

Until now, we only used that $2 \leqslant \rho \leqslant \nu/3$. If we impose the condition

$$\rho \leqslant \nu \min\left(\frac{2c_Q}{\gamma + 3\log_{689} Q}||(a^2+2a+2)\alpha||^2, \frac{2\kappa}{1+\gamma}\right),$$

then we have

$$S \ll \nu^{\omega(q)/2+1} Q^{\nu - \gamma\rho/2}.$$

We set

$$c_{q,\kappa} = \min\left(\frac{2\kappa}{1+\gamma}, \frac{2c_Q}{\gamma + 3\log_{689} Q}\right),$$

and choose $\rho := \nu \cdot \left\lfloor c_{q,\kappa} \left\|(a^2+2a+2)\alpha\right\|\right\rfloor$. If $\rho \geqslant 2$, then we just have shown the desired result (cf. (3.5)). If $\rho < 2$, then the estimate (3.5) holds trivially and we are done.

3.5 Proofs of Corollary 3.3 and Corollary 3.4

Corollary 3.3 and Corollary 3.4 can be deduced from Theorem 3.2. Before we prove them, we show an auxiliary lemma concerning κ-$\mathbb{Z}[i]$ sequences.

Lemma 3.17. *Let $b, d \in \mathbb{Z}[i]$ and $(\mathcal{D}_N)_{N\in\mathbb{N}}$ be a κ-$\mathbb{Z}[i]$ sequence. Then we have*

$$\#\left\{z \in \mathcal{D}_N : z^2 \equiv b \bmod d\right\} = \frac{\#\mathcal{D}_N}{|d|^2}\#\left\{z \in R_d : z^2 \equiv b \bmod d\right\} + O_d\left(N^{1-\kappa}\right).$$

Proof. We define for $z \in \mathbb{Z}[i]$ the numbers ω_z to be 1 if $z^2 \equiv b \bmod d$ and 0 otherwise. Then we can write

$$\begin{aligned}\#\left\{z \in \mathcal{D}_N : z^2 \equiv b \bmod d\right\} &= \frac{1}{|d|^2}\sum_{z\in\mathcal{D}_N}\sum_{r\in R_d}\omega_{z+r} \\ &\quad + \frac{1}{|d|^2}\left(|d|^2\sum_{z\in\mathcal{D}_N}\omega_z - \sum_{r\in R_d}\sum_{z\in\mathcal{D}_N}\omega_{z+r}\right),\end{aligned}$$

where we choose R_d to be a complete residue system modulo d with $|r| \ll |d|$ for all $r \in R_d$. The first term is equal to

$$\frac{\#\mathcal{D}_N}{|d|^2}\#\left\{z \in R_d : z^2 \equiv b \bmod d\right\}.$$

It remains to estimate the error term. We have

$$\begin{aligned}\left||d|^2\sum_{z\in\mathcal{D}_N}\omega_z - \sum_{r\in R_d}\sum_{z\in\mathcal{D}_N}\omega_{z+r}\right| &\ll \sum_{r\in R_d}\left|\sum_{z\in\mathcal{D}_N}\omega_z - \sum_{z\in\mathcal{D}_N}\omega_{z+r}\right| \\ &\ll \sum_{r\in R_d}\#\left\{\mathcal{D}_N\triangle(r+\mathcal{D}_N)\right\} \\ &\ll \sum_{|r|\leqslant R_d}|r|N^{1-\kappa} \ll |d|^3N^{1-\kappa}.\end{aligned}$$

□

Proof of Corollary 3.3. We have

$$\#\left\{z \in \mathcal{D}_N : s_q(z^2) \equiv b \bmod g\right\} = \sum_{z\in\mathcal{D}_N} \frac{1}{g} \sum_{0\leqslant j<g} \mathrm{e}\left(\frac{j}{g}(s_q(z^2)-b)\right).$$

Set $\widetilde{d} = (g, a^2+2a+2)$. From Proposition A.4 follows that

$$s_q(z^2) = s_q((z_1+iz_2)^2) \equiv z_1^2 - z_2^2 \pm 2(a+1)z_1z_2 \bmod \widetilde{d}.$$

If we put $g' = \frac{g}{\widetilde{d}}$, $J = \{kg' : 0 \leqslant k < \widetilde{d}\}$, $J' = \{0,\dots,g-1\} \setminus J = \{kg'+r : 0 \leqslant k < \widetilde{d}, 1 \leqslant r < g'\}$, then we have for $j = kg' \in J$,

$$\mathrm{e}\left(\frac{j}{g}s_q(z^2)\right) = \mathrm{e}\left(\frac{k}{\widetilde{d}}s_q(z^2)\right) = \mathrm{e}\left(\frac{k}{\widetilde{d}}\left(z_1^2 - z_2^2 \pm 2(a+1)z_1z_2\right)\right).$$

Hence,

$$\begin{aligned}&\sum_{z\in\mathcal{D}_N} \frac{1}{g} \sum_{j\in J} \mathrm{e}\left(\frac{j}{g}(s_q(z^2)-b)\right)\\ &\qquad = \sum_{z\in\mathcal{D}_N} \frac{1}{g} \sum_{0\leqslant k<\widetilde{d}} \mathrm{e}\left(\frac{k}{\widetilde{d}}\left(z_1^2 - z_2^2 \pm 2(a+1)z_1z_2 - b\right)\right)\\ &\qquad = \frac{\widetilde{d}}{g}\#\left\{z = z_1+iz_2 \in \mathcal{D}_N : z_1^2 - z_2^2 \pm 2(a+1)z_1z_2 \equiv b \bmod \widetilde{d}\right\}.\end{aligned}$$

Since $(\widetilde{d}, q-1) = ((g, |q-1|^2), q-1) = (g, q-1) = d$, we get by Lemma 2.22 and Lemma 3.17 that the last quantity is the same as

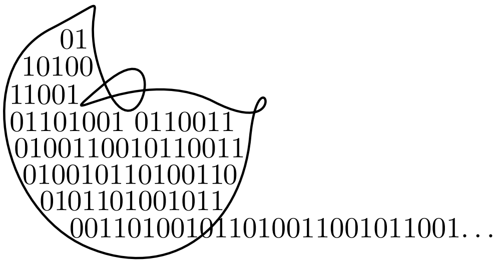
$$\frac{\widetilde{d}}{g}\#\left\{z \in \mathcal{D}_N : z^2 \equiv b \bmod d\right\} = \frac{\#\mathcal{D}_N}{g}\frac{\widetilde{d}}{|d|^2}Q(b,d) + O_{q,g}(N^{1-\kappa}).$$

Using the same considerations as in Lemma 2.22, we see that $\widetilde{d}/|d|^2 = 1$. The remaining sum (which comes from the set J') can be treated with Theorem 3.2 and one finally obtains Corollary 3.3 (see the proof of Theorem 3 in [MR09] for details). □

Proof of Corollary 3.4. If $\alpha \in \mathbb{Q}$, then the sequence $(\alpha s_q(z^2))_{z\in\mathbb{Z}[i]}$ takes modulo 1 only a finite number of values and is therefore not uniformly distributed modulo 1. If in return $\alpha \in \mathbb{R}\setminus\mathbb{Q}$, then for every $h \in \mathbb{Z}$ with $h \neq 0$ we have $(a^2+2a+2)h\alpha \in \mathbb{R}\setminus\mathbb{Q}$ and according to Theorem 3.2 (take for $\mathcal{D}_N$ the Gaussian integers lying in a disc with radius $\sqrt{N}$), the statement follows from Weyl's criterion (Theorem A.18). Note, that the specific ordering (if only the absolute value of the numbers increases monotonically) of the Gaussian integers is negligible. □

Chapter 4

Generalized Thue-Morse sequence in compact groups



The Thue-Morse sequence

$$(t_n)_{n\in\mathbb{N}} = (0110100110010110100101100110 1001\ldots)$$

has the property that the digits 0 and 1 appear with asymptotic frequency 1/2. As shown by Gelfond, this property persists for subsequences like linear progressions. It follows from Mauduit's and Rivat's solution of Gelfond's sum-of-digits problem of squares that the sequence $(t_{n^2})_{n\in\mathbb{N}}$ has the same property.

In this chapter we consider compact group generalizations $T(n)$ of the Thue-Morse sequence and prove that the subsequence $T(n^2)$ is uniformly distributed with respect to a measure ν that is absolutely continuous with respect to the Haar measure. The proof is based on a proper generalization of the Fourier based method of Mauduit and Rivat to group representations.

4.1 Introduction and main results

The Thue-Morse sequence $(t_n)_{n\in\mathbb{N}}$ has been rediscovered several times in the literature and there are various different definitions (see [Mau01]). For example, we have

$$t_n = s_2(n) \bmod 2,$$

where $s_2(n)$ denotes the number of 1's in the binary expansion of n. Alternatively, we can use recursive definitions like $t_0 = 0$, $t_{2k} = t_k$, $t_{2k+1} = 1 - t_k$ or identify it with a fixed point of the morphism $\mu : \{0,1\}^* \to \{0,1\}^*$ induced by $\mu(0) = 01$ and $\mu(1) = 10$ (see [Fog02]). Another quite non-standard definition is the following one: Define $u_0 = 0$, and for $j \geqslant 0$ set

$$u_{j+1} = \begin{cases} 0, & \text{if } \prod_{0 \leqslant i \leqslant j} \left(\frac{2i+1}{2i+2}\right)^{(-1)^{u_i}} > \frac{\sqrt{2}}{2}, \\ 1, & \text{otherwise.} \end{cases}$$

Then the sequence $(u_n)_{n\in\mathbb{N}}$ is the Thue-Morse sequence $(t_n)_{n\in\mathbb{N}}$ (see [AC85]). This sequence has many interesting properties. For example, it is cube-free (that is, there is no subword of the form www) and every subword w that occurs once appears infinitely often with bounded gaps (although it is non-periodic). It is also an automatic sequence (see [AS03] and Chapter 5). In any case, the binary expansion of n governs the behavior of t_n.

The purpose of this chapter is to establish a distribution result for the quadratic subsequence $T(n^2)$, where $T(n)$ is a generalized Thue-Morse sequence of the following type: Let H be a compact group that satisfies the Hausdorff separation axiom, $q \geqslant 2$, and $g_0, g_1, \dots, g_{q-1} \in H$ with $g_0 = e$ the identity element. Furthermore, let $G \leqslant H$ be the closure of the subgroup generated by $g_0, g_1, \dots, g_{q-1}$. Suppose that

$$\begin{aligned} n &= \varepsilon_{\ell-1}(n)q^{\ell-1} + \varepsilon_{\ell-2}(n)q^{\ell-2} + \cdots + \varepsilon_1(n)q + \varepsilon_0(n) \\ &= (\varepsilon_{\ell-1}(n)\varepsilon_{\ell-2}(n)\dots\varepsilon_1(n)\varepsilon_0(n))_q \end{aligned}$$

denotes the q-ary digital expansion of n and define

$$T(n) = g_{\varepsilon_0(n)} g_{\varepsilon_1(n)} \cdots g_{\varepsilon_{\ell-1}(n)}. \tag{4.1}$$

If $G = \mathbb{Z}/2\mathbb{Z}$ (with $+$ as the group operation), $q = 2$, and $g_0 = 0$, $g_1 = 1$, then $T(n) = s_2(n) \bmod 2 = t_n$. Thus, $T(n)$ is a proper generalization of the Thue-Morse sequence. Alternatively $T(n)$ can be seen as a completely q-multiplicative G-valued function which is defined by the property

$$T(j + qn) = T(j)T(n)$$

for $n \geqslant 0$ and $0 \leqslant j < q$. The sequence $T(n)$ is also an example of a chained sequence with a transition matrix that is not contractive (see [AL91] and [AL09]).

It is relatively easy to show (see Theorem 4.10) that the sequence $(T(n))_{n\geqslant 0}$ is uniformly distributed in G, that is, the normalized counting measure induced by $T(n)$, $n < N$, converges weakly to the (normalized) Haar measure μ on G:[1]

$$\frac{1}{N}\sum_{n=0}^{N-1} \delta_{T(n)} \to \mu.$$

Our main result deals with the question whether this remains true if $T(n)$ is replaced by the subsequence of squares $T(n^2)$. Actually, this sequence is not necessarily uniformly distributed. Nevertheless, there is always a measure ν such that $T(n^2)$ is ν-uniformly distributed.

[1] The symbol δ_x denotes the point measure concentrated at x.

Theorem 4.1. *Let $T(n)$ be defined by (4.1). Then there exists a positive integer m depending on $g_0 = e, g_1, \dots, g_{q-1}$ and q with $m \mid q-1$ such that the following holds.*

The group[2] *$U = \mathrm{cl}(\{T(mn) : n \geqslant 0\})$ is a normal subgroup of G of index m with cosets $g_u U = \mathrm{cl}(\{T(mn+u) : n \geqslant 0\})$, $0 \leqslant u < m$. With the help of these cosets we define*

$$\mathrm{d}\nu = \sum_{u=0}^{m-1} \mathbf{1}_{g_u U} \cdot Q(u,m)\,\mathrm{d}\mu,$$

where $Q(u,m) = \#\{0 \leqslant n < m : n^2 \equiv u \bmod m\}$ and μ denotes the Haar measure on G. Then the sequence $(T(n^2))_{n \geqslant 0}$ is ν-uniformly distributed in G, that is,

$$\frac{1}{N} \sum_{n=0}^{N-1} \delta_{T(n^2)} \to \nu.$$

Remark 4.2. The integer m that we will call *characteristic integer of $g_0, \dots, g_{q-1}$ and q* (see Section 4.2) is defined as the largest integer such that $m \mid q-1$ and such that there exists a one-dimensional representation D of G with

$$D(g_u) = \mathrm{e}\left(-\frac{u}{m}\right) \qquad \text{for all } u \in \{0, 1, \dots, q-1\}.$$

Note also that if $m = 1$ or $m = 2$ then $\nu = \mu$. Hence, if $m \leqslant 2$ then $(T(n^2))_{n \geqslant 0}$ is uniformly distributed in G. In particular if $q = 2$ or $q = 3$ then $m \leqslant 2$. Furthermore it is easy to observe that $\nu \neq \mu$ for $m > 2$, that is, $T(n^2)$ is not uniformly distributed in these cases.

Remark 4.3. Theorem 4.1 is a generalization of the results of Mauduit and Rivat [MR09]. Suppose first that $H = \mathbb{Z}/r\mathbb{Z}$ and $g_j = j \bmod r$, $0 \leqslant j < q$. Then $T(n) = s_q(n) \bmod r$ ($s_q(n) = \varepsilon_{\ell-1}(n) + \varepsilon_{\ell-2}(n) + \cdots + \varepsilon_1(n) + \varepsilon_0(n)$ denotes the q-ary sum-of-digits function) and Theorem 4.1 translates into Theorem 3 from [MR09] on the distribution of $s_q(n^2)$ modulo r (the characteristic integer is given by $(q-1, r)$, see also Chapter 5).

Similarly, if $H = \mathbb{R}/\mathbb{Z}$ and $g_j = \alpha j \bmod 1$, $0 \leqslant j < q$ for some irrational number α then $T(n) = \alpha s_q(n) \bmod 1$. In this case we have $G = H = \mathbb{R}/\mathbb{Z}$. Hence, Theorem 4.1 implies that $(\alpha s_q(n^2))_{n \in \mathbb{N}}$ is uniformly distributed modulo 1 (for example, one can use the fact that H is connected, see Remark 4.5); this is Theorem 2 from [MR09].

Remark 4.4. It is easy to derive some corollaries from Theorem 4.1. For example we have for all $0 \leqslant u < q$

$$\lim_{N \to \infty} \frac{1}{N} \#\{0 \leqslant n < N : T(n^2) \in g_u U\} = \frac{Q(u,m)}{m}.$$

A similar idea applies to compact homogeneous spaces X. Let H be the group acting on X and suppose that $g_0, \dots, g_{q-1}$ are chosen in a way that $G = H$. Then it follows that for every $x_0 \in X$ the sequence $x_n = T(n) \cdot x_0$ is uniformly distributed on X and the distribution behavior of x_{n^2} can be determined, too. For example, with the help of this approach one can construct uniformly distributed sequences on the sphere S^d.

[2] If $A \subseteq G$, then $\mathrm{cl}(A)$ denotes the topological closure of A in G

Remark 4.5. It follows from the proof of Theorem 4.1 that the Radon-Nikodym derivative $f(g) = Q(u,m)$ (for $g \in g_u U$) is continuous, which implies that G cannot be connected if $(T(n^2))_{n \geqslant 0}$ is not uniformly distributed. Conversely if G is connected then $(T(n^2))_{n \geqslant 0}$ is definitely uniformly distributed. Similarly, if the commutator subgroup of G coincides with G (i.e., G is a perfect group), then $(T(n^2))_{n \geqslant 0}$ is also uniformly distributed. Note that the commutator subgroup (the subgroup generated by the elements $xyx^{-1}y^{-1}$) is always a subgroup of U.

Remark 4.6. It would be also of interest to consider the subsequence $(T(p))$, where p runs over all primes. For example, an equidistribution result holds for t_p (see [MR10]). In order to handle this case one would need estimates of the form

$$\sum_{0 \leqslant h < q^\lambda} \|F_\lambda(h)\| \ll q^{\eta\lambda} \tag{4.2}$$

for some $\eta < 1/2$, where $F_\lambda(h)$ is the Fourier term defined in Section 4.2. By using the Cauchy-Schwarz inequality it follows directly that (4.2) holds for $\eta = 1/2$. However, it is not clear how to derive such a general estimate for $\eta < 1/2$. Actually, this is one of the key estimates in [MR10], where the sum-of-digits function of primes is discussed.

Remark 4.7. If $g_0 \neq e$, then the sequence $(T(n))_{n \geqslant 0}$ is not q-multiplicative any more. This would not be essential for the proof of the main theorem since it is possible to reduce the function $T(n)$ to the function $T_\lambda(n)$ defined in Section 4.3.2, which is "almost" completely q-multiplicative even if $g_0 \neq e$ (it satisfies $T_\lambda(j+qn) = g_j T_{\lambda-1}(n)$ for all $0 \leqslant j < q$ and $n \geqslant 0$).

However, the condition $g_0 = e$ is important for the proof of Lemma 4.9 and Lemma 4.13. It is only possible to avoid this condition if one assumes instead that the group G is equal to the closure of the subgroup generated by $g_i^{-1} g_j, 0 \leqslant i,j < q$ and that there exists no one-dimensional representation D satisfying

$$D(g_u) = \mathrm{e}(-tu) D(g_0)$$

for all $0 \leqslant u < q$ with $t(q-1) \in \mathbb{Z}$ and $D(g_0) \neq 1$. However, for the sake of brevity we use the assumption $g_0 = e$ in the main theorem, since this yields a considerably simpler presentation of the proof.

The proof of Theorem 4.1 is based on a proper generalization of the Fourier-based method of Mauduit and Rivat [MR10, MR09] to group representations. In Section 4.2 we use representation theory to prove uniform distribution of the sequence $(T(n))_{n \geqslant 0}$ and develop the theory to discuss the case of linear subsequences $(T(an + b))_{n \geqslant 0}$ (see Theorem 4.15). Although linear subsequences are not the main focus of this work the analysis of them is also useful for the analysis of the quadratic subsequence $(T(n^2))_{n \geqslant 0}$. Interestingly, the characteristic integer m appears there in a quite natural way. The technical part of the proof of Theorem 4.1 is presented in Sections 4.3 and 4.4, where we first establish some auxiliary results (like matrix generalizations of the techniques used in [MR09]) and then collect all necessary facts to complete the proof.

We will see in Chapter 5 an application of Theorem 4.1. In particular, we will deal with finite groups G in more detail and also show that there is a close relation of $T(n)$ to so-called automatic sequences. Actually, this kind of application was the main motivation of the present study.

4.2 Group representations

A unitary group representation D of a compact group is a continuous homomorphism $D : G \to U_d$ for some $d \geqslant 1$, where U_d denotes the group of unitary $d \times d$ matrices (over $\mathbb{C}$). A representation is irreducible if there is no proper subspace W of $\mathbb{C}^d$ with $D(x)W \subseteq W$ for all $x \in G$. The trivial representation that maps all elements to 1 is denoted by D_0 and the dimension d is called the dimension (or degree) of D.

Irreducible and unitary group representations can be used to prove uniform distribution of a sequence (x_n) in a compact group.

Lemma 4.8. *Let G be a compact group and ν a regular normed Borel measure in G. Then a sequence $(x_n)_{n\geqslant 0}$ is ν-uniformly distributed in G if and only if*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} D(x_n) = \int_G D \,\mathrm{d}\nu$$

holds for all irreducible and unitary representations D of G. In particular, $(x_n)_{n\geqslant 0}$ is uniformly distributed in G (with respect to the Haar measure μ) if and only if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} D(x_n) = 0$$

holds for all irreducible and unitary representations $D \neq D_0$.

Proof. A proof of this lemma can be found for example in [KN74, Theorem 1.3 and Section 4.3]. □

We will first present a proof that a sequence $T(n)$ of the form (4.1) is uniformly distributed in G. Let D be a representation of G and let

$$\Psi_D = \sum_{0\leqslant u<q} D(g_u)$$

denote the sum of the matrices $D(g_0), \ldots, D(g_{q-1})$.

We use the following notations for matrices. If A is a matrix, then A^H is the Hermitian transpose, $\rho(A)$ denotes the spectral radius and $\mathrm{tr}(A)$ the trace of A. We use[3] $\|\,.\,\|_2$ for the spectral norm ($\|A\|_2 = \sqrt{\rho(AA^H)}$) and $\|A\|_{\mathbb{F}}$ for the Frobenius norm (i.e., $\|A\|_{\mathbb{F}}^2 = \sum_{i,j} |a_{i\,j}|^2 = \mathrm{tr}(AA^H)$).

[3] We want to remark, that in this chapter $\|\cdot\|$ does not denote the function that maps $x \in \mathbb{R}$ to the distance to its nearest integer.

Lemma 4.9. *Let G be a compact group that is the closure of the subgroup generated by the elements $g_0, g_1, \ldots, g_{q-1}$, where $g_0 = e$. Suppose that $D \neq D_0$ is an irreducible and unitary representation of G. Then*

$$\|\Psi_D\|_2 < q. \tag{4.3}$$

Proof. Suppose that $\mathbf{y} \in \mathbb{C}^d$ is a non-zero vector. Then $\|D(x)\mathbf{y}\|_2 = \|\mathbf{y}\|_2$ for all $x \in G$, and consequently

$$\|\Psi_D \mathbf{y}\|_2 = \left\| \sum_{0 \leqslant u < q} D(g_u)\mathbf{y} \right\|_2 \leqslant \sum_{0 \leqslant u < q} \|D(g_u)\mathbf{y}\|_2 = q\,\|\mathbf{y}\|_2\,. \tag{4.4}$$

Hence $\|\Psi_D\|_2 \leqslant q$. Suppose now that $\|\Psi_D\|_2 = q$, that is, there exists a non-zero vector $\mathbf{y}$ with $\|\Psi_D \mathbf{y}\|_2 = q\,\|\mathbf{y}\|_2$. We have

$$\|\Psi_D \mathbf{y}\|_2^2 = \sum_{0 \leqslant u,v < q} \langle D(g_u)\mathbf{y}, D(g_v)\mathbf{y}\rangle = q^2 \langle \mathbf{y}, \mathbf{y}\rangle\,.$$

The Cauchy-Schwarz inequality implies

$$|\langle D(g_u)\mathbf{y}, D(g_v)\mathbf{y}\rangle| \leqslant \|D(g_u)\mathbf{y}\|_2\,\|D(g_v)\mathbf{y}\|_2 = \|\mathbf{y}\|_2^2\,. \tag{4.5}$$

Since $\langle D(g_0)\mathbf{y}, D(g_0)\mathbf{y}\rangle = \langle \mathbf{y}, \mathbf{y}\rangle$, we have that $\langle D(g_u)\mathbf{y}, D(g_v)\mathbf{y}\rangle = \langle \mathbf{y}, \mathbf{y}\rangle$ for all $0 \leqslant u, v < q$ and there has to be equality in (4.5). It follows that $D(g_u)\mathbf{y}$ and $D(g_v)\mathbf{y}$ have to be linear dependent. Since $D(g_0)\mathbf{y} = \mathbf{y}$, we obtain

$$D(g_u)\mathbf{y} = \mathbf{y} \tag{4.6}$$

for all $u = 0, 1, \ldots, q-1$. Consequently the one-dimensional space $W = \operatorname{span}(\mathbf{y})$ satisfies $D(x)W \subseteq W$ for all $x \in G$. (Recall that G is the closure of the subgroup generated by $g_0, \ldots, g_{q-1}$.) This contradicts the assumption that D is irreducible provided that the dimension of D is greater than or equal to two. Thus, for all irreducible representations of dimension $d \geqslant 2$ we actually have $\|\Psi_D\|_2 < q$.

Finally suppose that $d = 1$, that is, we are considering characters. Then (4.6) says that $D(g_u) = 1$ for all $u = 0, 1, \ldots, q-1$. Since G is the closure of the subgroup generated by the elements $g_0, g_1, \ldots, g_{q-1}$ this would imply $D(x) = 1$ for all $x \in G$ which contradicts the assumption $D \neq D_0$. □

With the help of Lemma 4.9 it is easy to prove that the sequence $(T(n))_{n \geqslant 0}$ is uniformly distributed in G:

Theorem 4.10. *Let $q \geqslant 2$ and $g_0, g_1, \ldots, g_{q-1}$ with $g_0 = e$ be elements of a compact group and G the closure of the subgroup generated by $g_0, g_1, \ldots, g_{q-1}$. Then the sequence $(T(n))_{n \geqslant 0}$ defined by (4.1) is uniformly distributed in G with respect to its Haar measure μ.*

Proof. Let $D \neq D_0$ be an irreducible and unitary representation of G and recall that $T(n) = g_{\varepsilon_0} g_{\varepsilon_1} \cdots g_{\varepsilon_{\ell-1}}$, where $\varepsilon_{\ell-1}\varepsilon_{\ell-2}\ldots\varepsilon_1\varepsilon_0$ denotes the q-ary digital expansion of n. Let $N \geqslant 1$ be defined by $N = \sum_{\nu=0}^{\lambda} n_\nu q^\nu$ with $n_\lambda \neq 0$. We begin with the following identity

$$\begin{aligned}\sum_{0\leqslant n<N} D(T(n)) &= \sum_{\nu=0}^{\lambda}\sum_{n=0}^{q^\nu-1}\sum_{\varepsilon_\nu=0}^{n_\nu-1} D(T(n+\varepsilon_\nu q^\nu + n_{\nu+1}q^{\nu+1}+\cdots+n_\lambda q^\lambda))\\ &= \sum_{\nu=0}^{\lambda}\left(\sum_{n=0}^{q^\nu-1} D(T(n))\right)\left(\sum_{\varepsilon_\nu=0}^{n_\nu-1} D(T(\varepsilon_\nu))\right) D(g_{n_{\nu+1}})\cdots D(g_{n_\lambda}).\end{aligned}$$

Since D is a unitary representation and the 2-norm is submultiplicative, we obtain

$$\begin{aligned}\left\|\sum_{0\leqslant n<N} D(T(n))\right\|_2 &\leqslant \sum_{\nu=0}^{\lambda}\sum_{\varepsilon_\nu=0}^{n_\nu-1} \|D(T(\varepsilon_\nu))\|_2 \left\|\sum_{n=0}^{q^\nu-1} D(T(n))\right\|_2\\ &\leqslant q\sum_{\nu=0}^{\lambda}\left\|\sum_{n=0}^{q^\nu-1} D(T(n))\right\|_2 .\end{aligned}$$

By induction it follows from the definition of $T(n)$ that $\sum_{n=0}^{q^\nu-1} D(T(n)) = (\Psi_D)^\nu$. Lemma 4.9 implies that we have $\|\Psi_D\|_2 = q^\sigma$, where $\sigma < 1$. We finally get[4]

$$\left\|\frac{1}{N}\sum_{0\leqslant n<N} D(T(n))\right\|_2 \leqslant \frac{q}{N}\sum_{\nu=0}^{\lambda}\|\Psi_D\|_2^\nu \ll \frac{q^{\sigma\lambda}}{N} \to 0$$

as N goes to infinity. Applying Lemma 4.8, this proves the theorem. □

It is an interesting problem to generalize Theorem 4.10 to special subsequences of $T(n)$, for example to linear subsequences $T(an+b)$ or (as it is the main goal of this chapter) to the subsequence $T(n^2)$ of squares. First we present a result for linear subsequences. Let D be an irreducible representation of G with degree d. In what follows, we need the function

$$\Psi_D(t) = \sum_{0\leqslant u<q} \mathrm{e}(tu)D(g_u),$$

and the *Fourier terms* (defined for $\lambda \geqslant 0$)

$$\begin{aligned}F_\lambda(h) &= \frac{1}{q^\lambda}\sum_{0\leqslant u<q^\lambda} e\left(-\frac{hu}{q^\lambda}\right) D(T(u))\\ &= \frac{1}{q^\lambda}\Psi_D\left(-\frac{h}{q^\lambda}\right)\Psi_D\left(-\frac{h}{q^{\lambda-1}}\right)\cdots\Psi_D\left(-\frac{h}{q}\right).\end{aligned}$$

[4] In this chapter, the implied constant in $f \ll g$ may at most depend on the base q, i.e., there exists a constant c only depending on q such that $|f| \leqslant cg$.

The product representation follows from the fact that

$$\begin{aligned} F_\lambda(h) &= \frac{1}{q^\lambda} \sum_{0\leqslant j<q} \sum_{0\leqslant u<q^{\lambda-1}} \mathrm{e}\left(-\frac{h(uq+j)}{q^\lambda}\right) D(T(uq+j)) \\ &= \frac{1}{q^\lambda} \sum_{0\leqslant j<q} \mathrm{e}\left(-\frac{hj}{q^\lambda}\right) D(g_j) \sum_{0\leqslant u<q^{\lambda-1}} \mathrm{e}\left(-\frac{hu}{q^{\lambda-1}}\right) D(T(u)) \\ &= \frac{1}{q}\, \Psi_D\left(-\frac{h}{q^\lambda}\right) F_{\lambda-1}(h), \end{aligned} \tag{4.7}$$

and that $F_0(h) = I_d$, where I_d is the identity matrix of dimension d. (Note, that $T(uq+j) = g_j T(u)$ for all $0 \leqslant j < q$ and that g_0 is the identity element.)

If the representation D is one-dimensional and satisfies

$$D(g_u) = \mathrm{e}\left(-\frac{ur}{q-1}\right)$$

for $0 \leqslant u < q$, then we see that D acts on the set $\{T(n) : n \geqslant 0\}$ in a special way. Indeed, since $s_q(n) \equiv n \bmod q-1$, we have

$$D(T(n)) = \mathrm{e}\left(-\frac{s_q(n)r}{q-1}\right) = \mathrm{e}\left(-\frac{nr}{q-1}\right).$$

Actually, these kinds of representations are crucial for the description of the distribution of $T(an+b)$ and $T(n^2)$.

Definition 4.11. Let m be the largest integer such that $m \mid q-1$ and such that there exists a one-dimensional representation D of G with

$$D(g_u) = \mathrm{e}\left(-\frac{u}{m}\right) \qquad \text{for all } u \in \{0,1,\dots,q-1\}. \tag{4.8}$$

We will call this integer *characteristic integer of* $g_0, \dots, g_{q-1}$ *and* q.

Observe that this characteristic integer m always exists since the trivial representation fulfills (4.8) with $m = 1$. The next Lemma collects some facts of this characteristic integer.

Lemma 4.12. *Let m be the characteristic integer of $g_0, \dots, g_{q-1}$ and q. Then there exist m representations $D_0, \dots, D_{m-1}$ of G with the following properties:*

(i) *Let $0 \leqslant k < m$. Then*

$$D_k(g_u) = \mathrm{e}\left(-\frac{k}{m}u\right) \qquad \textit{for all } u \in \{0,1,\dots,q-1\}.$$

(ii) *All other representations of G do not satisfy $D(g_1) = \mathrm{e}(-t)$ and $D(g_u) = D(g_1)^u$ for all $0 \leqslant u < q$ with $(q-1)\,t \in \mathbb{Z}$.*

(iii) *The kernel* $\ker D_1 = \{g \in G : D_1(g) = 1\}$ *is a normal subgroup of* G *and the index of* $\ker D_1$ *in* G *is equal to* m.

(iv) *The* m *cosets of* $\ker D_1$ *are given by*

$$g_v \ker D_1 = \operatorname{cl}\left(\{T(mn+v) : n \geqslant 0\}\right) \qquad \textit{for all } v \in \{0,1,\ldots,m-1\}.$$

Proof. Let D be a one-dimensional representation of G that satisfies (4.8) and set

$$D_k(g) = D(g)^k$$

for all $g \in G$ and for all $0 \leqslant k < m$. Then D_0 (consistent with the already defined notation) is the trivial representation and the representations D_k satisfy the relation stated in (i) for all $0 \leqslant k < m$. (Note, that the functions D_k are indeed representations coming from the (iterated) tensor product of the representation D.) Next, we show that there are no other representations of G with this property. Assume, that there exists a representation $\tilde{D} \neq D_k$ for all $0 \leqslant k < m$ such that

$$\tilde{D}(g_u) = \mathrm{e}\left(-\frac{r}{m'}h\right) \qquad \text{for all} \quad 0 \leqslant u < q,$$

and for some integers $m' \geqslant 1$, $r \geqslant 0$ with $(r,m') = 1$ and $m' \mid q-1$. Then $(m,m') < m'$ and there exist non-negative integers x and y such that $xm' + yrm \equiv (m,m') \bmod mm'$ (note, that $(rm,m') = (m,m')$). If we set

$$\chi(g) = D_1^x(g)\tilde{D}^y(g),$$

then χ is a representation satisfying

$$\chi(g_u) = \mathrm{e}\left(-u\left(\frac{x}{m} + \frac{yr}{m'}\right)\right) = \mathrm{e}\left(-u\frac{m'x + yrm}{mm'}\right) = \mathrm{e}\left(-\frac{u}{\bar{m}}\right),$$

for all $0 \leqslant u < q$, where $\bar{m} = \operatorname{lcm}(m,m')$. Since $\bar{m} = mm'/(m,m') > m$, this is impossible by the definition of m. Thus, we have shown (ii). The kernel of a representation is clearly a normal subgroup and the factor group $G/\ker D_1$ is isomorph to the image of D_1. Let $0 \leqslant v < m$. Then we have

$$D_1(T(mn+v)) = \mathrm{e}\left(-\frac{s_q(mn+v)}{m}\right) = \mathrm{e}\left(-\frac{mn+v}{m}\right) = \mathrm{e}\left(-\frac{v}{m}\right).$$

We see that

$$D_1\left(\operatorname{cl}\left(\{T(mn+v) : n \geqslant 0\}\right)\right) = \mathrm{e}\left(-\frac{v}{m}\right).$$

Since D_1 is continuous and $(T(n))_{n\geqslant 0}$ is dense in G (recall, that $(T(n))_{n\geqslant 0}$ is uniformly distributed in G), the family $\operatorname{cl}(\{T(mn+v) : n \geqslant 0\})$, $0 \leqslant v < m$ is a partition of G. We obtain that $G/\ker D_1$ is isomorph to the *m-th* roots of unity (and hence to $\mathbb{Z}/m\mathbb{Z}$). Moreover, we see that $\ker D_1 = \operatorname{cl}(\{T(mn) : n \geqslant 0\})$ and m is the index of $\ker D_1$ in G. Since

$$D_1(g_v^{-1}\{T(mn+v) : n \geqslant 0\}) = 1,$$

we finally have that $g_v \ker D_1 = \operatorname{cl}(\{T(mn+v) : n \geqslant 0\})$, $v = 0,\ldots,m-1$ are the m different cosets of $\ker D_1$. □

Lemma 4.13. *Let G be a compact group that is the closure of the subgroup generated by the elements $g_0 = e, g_1, \dots, g_{q-1}$. Suppose that $D \notin \{D_0, \dots, D_{m-1}\}$ is an irreducible and unitary representation of G. Then there exists a constant $c > 0$ such that*

$$\max_{h\in\mathbb{Z}} \|F_\lambda(h)\|_2 \ll q^{-c\lambda}.$$

Remark 4.14. If $D = D_k$ for some $0 \leqslant k < m$, then $|F_\lambda(h)| = q^{-\lambda}\varphi_{q^\lambda}(h/q^\lambda - k/m)$. (For the definition and the properties of φ_{q^λ} see Lemma A.6.) In particular, this implies that

$$\lim_{\lambda\to\infty} \max_{h\in\mathbb{Z}} |F_\lambda(h)| \neq 0,$$

and Lemma 4.13 cannot hold true in this case.

Proof. The proof is very similar to the proof of Lemma 4.9. We begin with considering the function $\Psi_D(t)$ for a fixed real number t. Let us first assume that D has dimension d greater than 1. It is clear that $\|\Psi_D(t)\|_2 \leqslant q$. Moreover, if $\|\Psi_D(t)\|_2 = q$ then there exists a non-zero vector $\mathbf{y}$ with $\|\Psi_D(t)\mathbf{y}\|_2 = q\,\|\mathbf{y}\|_2$. By the same reasoning as in Lemma 4.9, we obtain that

$$e(ut)D(g_u)\mathbf{y} = D(g_0)\mathbf{y} = \mathbf{y} \tag{4.9}$$

for all $u = 0, 1, \dots, q-1$. This contradicts the assumption that D is irreducible (we assumed that $d \geqslant 2$). Thus, for all irreducible representations of dimension $d \geqslant 2$ we have $\|\Psi_D(t)\|_2 < q$. If the dimension of D is equal to 1, then $\|\Psi_D(t)\|_2 = q$ means $\mathrm{e}(ut)D(g_u) = 1$ for all $u = 0, 1, \dots, q-1$ which is equivalent to $D(g_1) = \mathrm{e}(-t)$ and $D(g_u) = D(g_1)^u$. Hence, if this is not true, then we also get $\|\Psi_D(t)\|_2 < q$. If $D(g_1) = \mathrm{e}(-t)$ and $D(g_u) = D(g_1)^u$, then $(q-1)t \notin \mathbb{Z}$ (recall that $D \notin \{D_0, \dots, D_{m-1}\}$) and we obtain

$$\|\Psi_D(t)\Psi_D(qt)\|_2 < q^2.$$

Indeed, as in the considerations above, the condition $\|\Psi_D(t)\Psi_D(qt)\|_2 = q^2$ would imply $D(g_1) = e(-t) = e(-qt)$ which contradicts the assumption $(q-1)t \notin \mathbb{Z}$.

Now we can finish the proof of Lemma 4.13. Since the spectral norm $\|\Psi_D(t)\|_2$ is a submultiplicative norm and a continuous function in t, we obtain

$$\sup_{t\in\mathbb{R}} \|\Psi_D(t)\Psi_D(qt)\|_2 < q^2.$$

Using the product representation of F_λ (see (4.7)), this implies that there exists a constant $c > 0$ such that

$$\max_{h\in\mathbb{Z}} \|F_\lambda(h)\|_2 \ll q^{-c\lambda}.$$

□

Theorem 4.15. *Let $q \geqslant 2$, $a \geqslant 1$ and $b \geqslant 0$ be integers and m the characteristic integer of $g_0, \dots, g_{q-1}$ and q. The sequence $(T(an+b))_{n\geqslant 0}$ is uniformly distributed in G (with respect to the Haar measure) if and only if $(a, m) = 1$.*

If $(a,m) > 1$, then there exists a normal subgroup U (with index (a,m) in G) such that $(T(an+b))_{n\geqslant 0}$ is ν-uniformly distributed on a coset of U, where ν is the (translated) Haar measure of U.

Proof. Let D be an irreducible and unitary representation of G with $D \neq D_k$ for all $0 \leqslant k < m$. Furthermore, let the integers λ and β be defined by $q^{\lambda-1} \leqslant N < q^{\lambda}$ and $a \leqslant q^{\beta-1}$. Then we have (for sufficiently large N) that $aN+b < q^{\lambda+\beta}$. We can write

$$\begin{aligned}\sum_{0\leqslant n<N} & D(T(an+b)) \\ &= \sum_{0\leqslant u<q^{\nu+\beta}} \sum_{0\leqslant n<N} D(T(u)) \cdot \frac{1}{q^{\lambda+\beta}} \sum_{0\leqslant h<q^{\lambda+\beta}} \mathrm{e}\left(\frac{h(an+b-u)}{q^{\lambda+\beta}}\right) \\ &= \sum_{0\leqslant h<q^{\lambda+\beta}} F_{\lambda+\beta}(h) \sum_{0\leqslant n<N} \mathrm{e}\left(\frac{h(an+b)}{q^{\lambda+\beta}}\right).\end{aligned}$$

The exponential sum can be easily calculated and we obtain

$$\left\| \sum_{0\leqslant n<N} D(T(an+b)) \right\|_2 \ll \sum_{0\leqslant h<q^{\lambda+\beta}} \|F_{\lambda+\beta}(h)\|_2 \cdot \min\left(N, \frac{1}{\left|\sin\frac{\pi ha}{q^{\lambda+\beta}}\right|}\right).$$

Since $D \neq D_k$, Lemma 4.13 implies that $\|F_{\lambda+\beta}(h)\|_2 \ll q^{-c(\lambda+\beta)}$ for some $c > 0$. We get

$$\left\| \sum_{0\leqslant n<N} D(T(an+b)) \right\|_2 \ll q^{-c(\lambda+\beta)} \sum_{0\leqslant h<q^{\lambda+\beta}} \min\left(N, \frac{1}{\left|\sin\frac{\pi ha}{q^{\lambda+\beta}}\right|}\right).$$

For the sum in the last expression, one can apply [MR10, Lemma 6] to obtain

$$\begin{aligned}\left\| \sum_{0\leqslant n<N} D(T(an+b)) \right\|_2 &\ll q^{-c(\lambda+\beta)} \left(N + (\lambda+\beta)q^{\lambda+\beta}\right) \\ &\ll \lambda q^{(1-c)\lambda} \ll N^{1-\sigma}\end{aligned}$$

with an appropriately chosen constant $\sigma > 0$. Thus, we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} D(T(an+b)) = 0.$$

Next, we consider the representations D_k, $0 \leqslant k < m$. We can use the fact that $D(T(n)) = \mathrm{e}(-nk/m)$ for all $n \geqslant 0$ and obtain

$$\frac{1}{N}\sum_{n=0}^{N-1} D_k(T(an+b)) = \frac{1}{N}\sum_{n=0}^{N-1} \mathrm{e}\left(-\frac{k(an+b)}{m}\right) = \mathrm{e}\left(-\frac{kb}{m}\right) \frac{1}{N}\sum_{n=0}^{N-1} \mathrm{e}\left(-\frac{kan}{m}\right).$$

Set $d = m/(a,m)$. If $k \equiv 0 \bmod d$ (this is equivalent to $m \mid ak$), then we have

$$\frac{1}{N}\sum_{n=0}^{N-1} \mathrm{e}\left(-\frac{kan}{m}\right) = 1.$$

If $k \not\equiv 0 \bmod d$, we obtain

$$\frac{1}{N}\left|\sum_{n=0}^{N-1} D_k(T(an+b))\right| = \frac{1}{N}\left|\sum_{n=0}^{N-1} \mathrm{e}\left(-\frac{kan}{m}\right)\right| = \frac{1}{N}\left|\frac{\sin(\pi Nka/m)}{\sin(\pi ka/m)}\right| \to 0,$$

as n goes to infinity. Thus we obtain for $0 \leqslant k < m$ that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} D_k(T(an+b)) = \begin{cases} \mathrm{e}\left(-\frac{kb}{m}\right), & \text{if } k \equiv 0 \bmod d, \\ 0, & \text{otherwise.} \end{cases}$$

Note, that $d = m$ if and only if $(a,m) = 1$. Thus, we just have shown that all irreducible and unitary representations satisfy the necessary and sufficient condition for $(T(an+b))_{n\geqslant 0}$ to be uniformly distributed in G if and only if $(a,m) = 1$ (see Lemma 4.8).

In order to complete the proof of Theorem 4.15, we have to deal with the case $(a,m) > 1$. We define the function f for all $g \in G$ by

$$f(g) = 1 + \mathrm{e}\left(\frac{db}{m}\right) D_d(g) + \mathrm{e}\left(\frac{2db}{m}\right) D_{2d}(g) + \cdots + \mathrm{e}\left(\frac{(m-d)b}{m}\right) D_{m-d}(g).$$

It is a real-valued positive function on G. To see this, choose an element $g \in \{T((m,a)n+b') : n \geqslant 0\}$, where $0 \leqslant b' < (m,a)$. We can write

$$\begin{aligned} f(g) &= \sum_{\ell=0}^{(m,a)-1} \mathrm{e}\left(\frac{\ell db}{m}\right) D_{\ell d}(g) \\ &= \sum_{\ell=0}^{(m,a)-1} \mathrm{e}\left(\frac{\ell db}{m} - \frac{\ell d((m,a)n+b')}{m}\right) \\ &= \sum_{k=0}^{(a,m)-1} \mathrm{e}\left(\frac{\ell(b-b')}{(a,m)}\right). \end{aligned}$$

Since $(T(n))_{n\geqslant 0}$ is uniformly distributed in G with respect to the Haar measure, we have that $(T(n))_{n\geqslant 0}$ is dense in G. Set $U_1 = \mathrm{cl}(\{T((m,a)n+b') : n \geqslant 0,\ 0 \leqslant b' < (a,m),\ b' \equiv b \bmod (m,a)\})$ and $U_2 = \mathrm{cl}(\{T((m,a)n+b') : n \geqslant 0,\ 0 \leqslant b' < (a,m),\ b' \not\equiv b \bmod (m,a)\})$. Then we have $G = U_1 \cup U_2$ and

$$f(g) = \begin{cases} (a,m), & \text{if } g \in U_1, \\ 0, & \text{otherwise.} \end{cases}$$

This proves the claim that f is positive. (Moreover, since f is continuous, we see that the group has to have more than one component in this case.) Using this function, we define the measure ν by[5]

$$\mathrm{d}\nu = f\mathrm{d}\mu.$$

In what follows, we show that the sequence $(T(an+b))_{n\geqslant 0}$ is ν-uniformly distributed in G. Let us consider a complete set of pairwise inequivalent irreducible unitary representations D^α, $\alpha \in \mathcal{A}$, where $\mathcal{A}$ is some index set. Put $e_{ij}^\alpha(g) = \sqrt{n_\alpha} d_{ij}^\alpha(g)$, where $D^\alpha = (d_{ij})_{1\leqslant i,j\leqslant n_\alpha}$. It follows from representation theory that the set $\{e_{ij}^\alpha\}$ forms a complete orthonormal system in the Hilbert space $L^2(G)$ with the scalar product $\langle f, g\rangle = \int_G f\overline{g}\mathrm{d}\mu$. We obtain for D_k, $k \equiv 0 \bmod d$ that

$$\int_G D_k f \mathrm{d}\mu = \sum_{\ell=0}^{(m,a)-1} \mathrm{e}\left(\frac{\ell db}{m}\right) \langle D_k, \overline{D_{\ell d}}\rangle = \mathrm{e}\left(-\frac{kb}{m}\right).$$

For all other representations $D^\alpha = (d_{ij}^\alpha)_{1\leqslant i,j\leqslant n_\alpha}$, we get

$$\int_G d_{ij}^\alpha f \mathrm{d}\mu = \sum_{\ell=0}^{(m,a)-1} \mathrm{e}\left(\frac{\ell db}{m}\right) \langle d_{ij}^\alpha, \overline{D_{\ell d}}\rangle = 0.$$

Lemma 4.8 implies that $(T(an+b))_{n\geqslant 0}$ is ν-uniformly distributed in G. If we set $U := \ker D_d$, then U is a normal subgroup of G (with index (m,a) in G). Similar to the proof of Lemma 4.12, one can show that

$$U = \mathrm{cl}\left(\{T((m,a)n) : n \geqslant 0\}\right)$$

and $U_1 = T(b)^{-1}U$, that is, U_1 is a coset of U. Since the support of ν is U_1, we have that $(T(an+b))_{n\geqslant 0}$ is dense in U_1. If we define the measure $\tilde{\nu}$ on U_1 by

$$\tilde{\nu}(B) = \int_B 1 \,\mathrm{d}\nu$$

for all Borel-sets B in U_1, we have that $(T(an+b))_{n\geqslant 0}$ is $\tilde{\nu}$-uniformly distributed in U_1. Moreover, $\tilde{\nu}$ is the translated normed Haar measure on U. Indeed, if we set

$$\tilde{\mu}(A) := \int_{T(b)^{-1}A} 1 \,\mathrm{d}\tilde{\nu}$$

for all Borel-sets A in U, then $\tilde{\mu}$ is translation invariant on U. This finishes the proof of Theorem 4.15. □

4.3 Auxiliary results

4.3.1 Van der Corput-type inequalities

We begin with two van der Corput-type inequality for matrices which enable us to "truncate" the sequence $T(n)$ twice (see Section 4.3.2 and Section 4.4). The presented

[5] That is, $\nu(A) = \int_A f\mathrm{d}\mu$ for all Borel-sets A.

lemmas are inspired by [MR09, Lemme 15 and 17]. We want to remark that van der Corput-type inequalities for matrices have been already considered by Hlawka [Hla55] (see also [KN74, Chapter 4.2]).

Lemma 4.16. *Let N and B be positive integers satisfying $N \leqslant B$. Furthermore let $Z(n) \in \mathbb{C}^{d\times d}$ for all $n \in \mathbb{Z}$ satisfying $\|Z(n)\|_{\mathbb{F}} \leqslant f$. Then we have for any real number $R \geqslant 1$,*

$$\left\|\sum_{0\leqslant n<N} Z(n)\right\|_{\mathbb{F}} \leqslant \left(\frac{\sqrt{d}N}{R}\sum_{|r|<R}\left(1-\frac{|r|}{R}\right)\left\|\sum_{0\leqslant n,n+r\leqslant B} Z(n+r)Z(n)^H\right\|_{\mathbb{F}}\right)^{1/2} + \frac{f}{2}R.$$

Proof. We take for convenience $Z(n) = 0$ (the $d \times d$ matrix consisting only of zeros) if $n \notin [0, B]$. We can write

$$\begin{aligned}\Bigg\|R\sum_{0\leqslant n<N} Z(n) - \sum_{-R/2<r\leqslant R/2}\ \sum_{0\leqslant n<N} Z(n+r)\Bigg\|_{\mathbb{F}} &\leqslant \sum_{-R/2<r\leqslant R/2}\left\|\sum_{0\leqslant n<N} Z(n) - \sum_{0\leqslant n<N} Z(n+r)\right\|_{\mathbb{F}}\\ &\leqslant \sum_{-R/2<r\leqslant R/2} 2f|r| \leqslant \frac{f}{2}R^2.\end{aligned}$$

Thus, we have

$$\left\|\sum_{0\leqslant n<N} Z(n)\right\|_{\mathbb{F}} \leqslant \frac{1}{R}\sum_{0\leqslant n<N}\left\|\sum_{-R/2<r\leqslant R/2} Z(n+r)\right\|_{\mathbb{F}} + \frac{f}{2}R.$$

Using the Cauchy-Schwarz inequality, we get

$$\left(\sum_{0\leqslant n<N}\left\|\sum_{-R/2<r\leqslant R/2} Z(n+r)\right\|_{\mathbb{F}}\right)^2 \leqslant N\sum_{0\leqslant n<N}\left\|\sum_{-R/2<r\leqslant R/2} Z(n+r)\right\|_{\mathbb{F}}^2. \quad (4.10)$$

Using the definition of the Frobenius norm and the fact that the trace is linear, the right-hand side of (4.10) is the same as

$$N\sum_{-R/2<r_1\leqslant R/2}\ \sum_{-R/2<r_2\leqslant R/2}\operatorname{tr}\left(\sum_{0\leqslant n\leqslant B} Z(n+r_1)\,Z(n+r_2)^H\right).$$

Changing the index of summation, we obtain that the last expression equals

$$N\sum_{-R<r<R}(R-|r|)\operatorname{tr}\left(\sum_{0\leqslant n\leqslant B} Z(n+r)\,Z(n)^H\right).$$

We have for all matrices $A = (a_{ij})_{1 \leqslant i,j \leqslant d}$ (using the Cauchy-Schwarz inequality)

$$|\operatorname{tr}(A)| \leqslant \sum_{1 \leqslant i \leqslant d} |a_{ii}| \leqslant \sqrt{d} \sqrt{\sum_{1 \leqslant i \leqslant d} |a_{ii}|^2} \leqslant \sqrt{d} \sqrt{\sum_{1 \leqslant i,j \leqslant d} |a_{ij}|^2} = \sqrt{d}\, \|A\|_{\mathbb{F}}\,, \tag{4.11}$$

and the statement of the lemma follows. □

Lemma 4.17. *Let N be a positive integer and $Z(n) \in \mathbb{C}^{d \times d}$ for all $0 \leqslant n \leqslant N$. Then we have for any real number $S \geqslant 1$ and any integer $k \geqslant 1$ the estimate*

$$\left\| \sum_{0 \leqslant n \leqslant N} Z(n) \right\|_{\mathbb{F}}^2 \leqslant \frac{N + k(S-1) + 1}{S} \sum_{|s| < S} \left(1 - \frac{|s|}{S}\right) \sum_{0 \leqslant n, n+ks \leqslant N} \operatorname{tr}\left(Z(n+ks) Z(n)^H\right).$$

Proof. Again, we take for convenience $Z(n) = 0$ (the $d \times d$ matrix consisting only of zeros) if $n \notin [0, N]$. Then we can write

$$S \sum_{n \in \mathbb{Z}} Z(n) = \sum_{n \in \mathbb{Z}} \sum_{0 \leqslant s < S} Z(n + ks).$$

If the last sum is not zero, then n satisfies $-k(S-1) \leqslant n \leqslant N$ and there are at most $N + k(S-1) + 1$ such values for n. Hence, applying the Cauchy-Schwarz inequality and changing the summation index yields (cf. Lemma 4.16)

$$\begin{aligned} S^2 \left\| \sum_{n \in \mathbb{Z}} Z(n) \right\|_{\mathbb{F}}^2 &\leqslant (N + k(S-1) + 1) \sum_{n \in \mathbb{Z}} \left\| \sum_{0 \leqslant s < S} Z(n + ks) \right\|_{\mathbb{F}}^2 \\ &\leqslant (N + k(S-1) + 1) \sum_{0 \leqslant s_1 < S} \sum_{0 \leqslant s_2 < S} \sum_{n \in \mathbb{Z}} \operatorname{tr}\left(Z(n + ks_1) Z(n + ks_2)^H\right) \\ &\leqslant (N + k(S-1) + 1) \sum_{|s| < S} (S - |s|) \sum_{n \in \mathbb{Z}} \operatorname{tr}\left(Z(n + ks) Z(n)^H\right). \end{aligned}$$

This proves the desired result. □

4.3.2 Fourier transform

Before we consider the Fourier transform, we recall the general setting. The compact group G is the closure of the subgroup generated by the elements $g_0, g_1, \ldots, g_{q-1}$ with $g_0 = e$ and D is an irreducible and unitary representation of G.

We start with defining truncated versions of the sequence $T(n)$. Let $\lambda \geqslant 1$ and $\mu < \lambda$ be positive integers. Set

$$T_\lambda(n) = g_{\varepsilon_0(n)} g_{\varepsilon_1(n)} \cdots g_{\varepsilon_{\lambda-1}(n)},$$

and

$$T_{\mu,\lambda}(n) = g_{\varepsilon_\mu(n)} g_{\varepsilon_{\mu+1}(n)} \cdots g_{\varepsilon_{\lambda-1}(n)},$$

where $\varepsilon_0(n), \dots, \varepsilon_{\lambda-1}(n)$ are the λ lower placed digits in the q-ary digital expansion (with possibly leading zeros) of n. Recall, that the function F_λ (which depends on D) equals

$$F_\lambda(h) = \frac{1}{q^\lambda} \sum_{0 \leqslant u < q^\lambda} \mathrm{e}\left(-\frac{hu}{q^\lambda}\right) D(T_\lambda(u)).$$

Additionally, we set

$$F_{\mu,\lambda}(h) := \frac{1}{q^\lambda} \sum_{0 \leqslant u < q^\lambda} \mathrm{e}\left(-\frac{hu}{q^\lambda}\right) D(T_{\mu,\lambda}(u)).$$

Before we start our treatment on these Fourier terms, we recall some definitions which we have already seen in the previous chapters or which are given in the appendix. The function φ_q is defined for all $q \geqslant 2$ by

$$\varphi_q(t) = \left| \sum_{0 \leqslant u < q} \mathrm{e}(ut) \right| = \begin{cases} \frac{|\sin \pi q t|}{|\sin \pi t|}, & \text{if } t \in \mathbb{R} \setminus \mathbb{Z} \\ q, & \text{if } t \in \mathbb{Z}. \end{cases}$$

Furthermore, the function

$$\Phi(q) = \max_{t \in \mathbb{R}} \frac{1}{q} \sum_{0 \leqslant r < q} \varphi_q\left(t + \frac{r}{q}\right),$$

satisfies $\Phi(q) \ll \log q$.

We start with showing a result on the second order average of $\|F_\lambda(h)\|_\mathbb{F}$ before we discuss the function $F_{\mu,\lambda}(h)$ in detail. The results we obtain here are very similar to the results in [MR09, Section 4.2]. Nevertheless, the proofs use many techniques from matrix theory. In what follows, the representation D has dimension $d \geqslant 1$.

Lemma 4.18. *Let $q \geqslant 2$, $a \in \mathbb{Z}$ and $0 \leqslant \delta \leqslant \lambda$. Then we have*

$$\sum_{\substack{0 \leqslant h < q^\lambda \\ h \equiv a \bmod q^\delta}} \|F_\lambda(h)\|_\mathbb{F}^2 = \|F_\delta(a)\|_\mathbb{F}^2. \tag{4.12}$$

Proof. If $\delta = \lambda$, the assertion is trivial. Hence, we assume that $\delta < \lambda$. Let us denote the left-hand side of (4.12) by S. Then we can write

$$\begin{aligned} S &= \sum_{\substack{0 \leqslant h < q^{\lambda-1} \\ h \equiv a \bmod q^\delta}} \sum_{0 \leqslant r < q} \left\| F_\lambda\left(h + r q^{\lambda-1}\right) \right\|_\mathbb{F}^2 \\ &= \sum_{\substack{0 \leqslant h < q^{\lambda-1} \\ h \equiv a \bmod q^\delta}} \sum_{0 \leqslant r < q} \mathrm{tr}\left(F_\lambda\left(h + r q^{\lambda-1}\right) F_\lambda\left(h + r q^{\lambda-1}\right)^H \right), \end{aligned}$$

which follows from the definition of the Frobenius norm. Using the recursive description of F_λ (see (4.7)), we obtain that this is the same as

$$\sum_{\substack{0\leqslant h<q^{\lambda-1}\\ h\equiv a \bmod q^\delta}} \sum_{0\leqslant r<q} \operatorname{tr}\left(\frac{1}{q^2}\Psi_D\left(-\frac{h+rq^{\lambda-1}}{q^\lambda}\right) F_{\lambda-1}(h)F_{\lambda-1}(h)^H \Psi_D\left(-\frac{h+rq^{\lambda-1}}{q^\lambda}\right)^H\right).$$

The trace is a linear operator with the property, that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ for two matrices A and B. Thus, we get that S is given by

$$\sum_{\substack{0\leqslant h<q^{\lambda-1}\\ h\equiv a \bmod q^\delta}} \operatorname{tr}\left(\frac{1}{q^2}\sum_{0\leqslant r<q} \Psi_D\left(-\frac{h+rq^{\lambda-1}}{q^\lambda}\right)^H \Psi_D\left(-\frac{h+rq^{\lambda-1}}{q^\lambda}\right) F_{\lambda-1}(h)F_{\lambda-1}(h)^H\right). \tag{4.13}$$

Next we claim that for every $t \in \mathbb{R}$ we have

$$\frac{1}{q^2}\sum_{0\leqslant r<q} \Psi_D\left(-t-\frac{r}{q}\right)^H \Psi_D\left(-t-\frac{r}{q}\right) = I_d. \tag{4.14}$$

Indeed, this holds true since the left-hand side of (4.14) is equal to

$$\begin{aligned}
&\frac{1}{q^2}\sum_{0\leqslant r<q}\sum_{0\leqslant u<q} \mathrm{e}\left(tu+\frac{ru}{q}\right) D(g_u)^H \sum_{0\leqslant v<q} \mathrm{e}\left(-tv-\frac{rv}{q}\right) D(g_v)\\
&\qquad = \frac{1}{q}\sum_{0\leqslant u<q}\sum_{0\leqslant v<q} \mathrm{e}\left(t(u-v)\right) D(g_u)^H D(g_v) \frac{1}{q}\sum_{0\leqslant r<q} \mathrm{e}\left(r\frac{u-v}{q}\right)\\
&\qquad = \frac{1}{q}\sum_{0\leqslant u<q} D(g_u)^H D(g_u),
\end{aligned}$$

and $D(g_u)^H D(g_u) = I_d$ for every $0 \leqslant u < q$. Using this result in (4.13), we obtain

$$\begin{aligned}
S &= \sum_{\substack{0\leqslant h<q^{\lambda-1}\\ h\equiv a \bmod q^\delta}} \operatorname{tr}\left(F_{\lambda-1}(h)F_{\lambda-1}(h)^H\right)\\
&= \sum_{\substack{0\leqslant h<q^{\lambda-1}\\ h\equiv a \bmod q^\delta}} \left\|F_{\lambda-1}(h)\right\|_{\mathbb{F}}^2 .
\end{aligned}$$

Applying this relation $\lambda - \delta$ times finally yields the desired result. □

Lemma 4.19. *Let $q \geqslant 2$ and $1 \leqslant \mu < \lambda$ be integers. Then we have*

$$\|F_{\mu,\lambda}(h)\| = \|F_{\lambda-\mu}(h)\| \, q^{-\mu}\varphi_{q^\mu}\left(hq^{-\lambda}\right), \tag{4.15}$$

where $\|\cdot\|$ is an arbitrary norm on $\mathbb{C}^{d\times d}$.

Proof. Since $T_{\mu,\lambda}(uq^\mu + v) = T_{\lambda-\mu}(u)$ for $0 \leqslant u < q^{\lambda-\mu}$ and $0 \leqslant v < q^\mu$, we get

$$\begin{aligned} F_{\mu,\lambda}(h) &= q^{-\lambda} \sum_{0 \leqslant u < q^{\lambda-\mu}} \sum_{0 \leqslant v < q^\mu} \mathrm{e}\left(-\frac{h(uq^\mu + v)}{q^\lambda}\right) D(T_{\mu,\lambda}(uq^\mu + v)) \\ &= q^{-(\lambda-\mu)} \sum_{0 \leqslant u < q^{\lambda-\mu}} \mathrm{e}\left(-\frac{hu}{q^{\lambda-\mu}}\right) D(T_{\lambda-\mu}(u)) q^{-\mu} \sum_{0 \leqslant v < q^\mu} \mathrm{e}\left(-\frac{hv}{q^\lambda}\right). \end{aligned}$$

This already implies the desired result. □

Lemma 4.20. *Let $q \geqslant 2$ and $1 \leqslant \mu < \lambda$ be integers. Then we have*

$$\sum_{0 \leqslant h < q^\lambda} \left\| F_{\mu,\lambda}(h) \right\|_2 \leqslant \Phi(q^\mu) q^{\lambda-\mu}. \tag{4.16}$$

Proof. In a first step, we can write

$$\begin{aligned} \sum_{0 \leqslant h < q^\lambda} \left\| F_{\mu,\lambda}(h) \right\|_2 &= \sum_{0 \leqslant u < q^{\lambda-\mu}} \sum_{0 \leqslant v < q^\mu} \left\| F_{\lambda-\mu}(u + vq^{\lambda-\mu}) \right\|_2 q^{-\mu} \varphi_{q^\mu}\left(\frac{u + vq^{\lambda-\mu}}{q^\lambda}\right) \\ &= \sum_{0 \leqslant u < q^{\lambda-\mu}} \left\| F_{\lambda-\mu}(u) \right\|_2 q^{-\mu} \sum_{0 \leqslant v < q^\mu} \varphi_{q^\mu}\left(\frac{u}{q^\lambda} + \frac{v}{q^\mu}\right). \end{aligned}$$

Next, note that $\|F_\lambda(h)\|_2 \leqslant 1$ since D is a unitary representation. This observation finally implies the result. □

Lemma 4.21. *Let $q \geqslant 2$, $a \in \mathbb{Z}$ and $1 \leqslant \lambda - \mu \leqslant \delta \leqslant \lambda$ be positive integers. Then we have*

$$\sum_{\substack{0 \leqslant h < q^\lambda \\ h \equiv a \bmod q^\delta}} \left\| F_{\mu,\lambda}(h) \right\| \leqslant \Phi\left(q^{\lambda-\delta}\right) q^{-\mu+\lambda-\delta} \varphi_{q^{\mu-\lambda+\delta}}\left(aq^{-\delta}\right) \left\| F_{\lambda-\mu}(a) \right\|,$$

where $\|\cdot\|$ is an arbitrary norm on $\mathbb{C}^{d\times d}$.

Proof. Since by assumption $\lambda - \mu \leqslant \delta$, we have $F_{\lambda-\mu}(a + \ell q^\delta) = F_{\lambda-\mu}(a)$. We obtain

$$\begin{aligned} \sum_{\substack{0 \leqslant h < q^\lambda \\ h \equiv a \bmod q^\delta}} \left\| F_{\mu,\lambda}(h) \right\| &= \sum_{0 \leqslant \ell < q^{\lambda-\delta}} \left\| F_{\lambda-\mu}\left(a + \ell q^\delta\right) \right\| q^{-\mu} \varphi_{q^\mu}\left(\frac{a + \ell q^\delta}{q^\lambda}\right) \\ &= \left\| F_{\lambda-\mu}(a) \right\| q^{-\mu} \sum_{0 \leqslant \ell < q^{\lambda-\delta}} \varphi_{q^\mu}\left(\frac{a}{q^\lambda} + \frac{\ell}{q^{\lambda-\delta}}\right). \end{aligned}$$

A short calculation shows (cf. [MR09, Proof of Lemma 12]), that

$$q^{-\mu} \sum_{0 \leqslant \ell < q^{\lambda-\delta}} \varphi_{q^\mu}\left(\frac{a}{q^\lambda} + \frac{\ell}{q^{\lambda-\delta}}\right) \leqslant \Phi\left(q^{\lambda-\delta}\right) q^{-\mu+\lambda-\delta} \varphi_{q^{\mu-\lambda+\delta}}\left(aq^{-\delta}\right).$$

This proves Lemma 4.21. □

Lemma 4.22. *Let $q \geqslant 2$, $a \in \mathbb{Z}$, $\ell \in \mathbb{Z}$, $1 \leqslant \mu < \lambda$ and $0 \leqslant \delta \leqslant \lambda - \mu$. Then we have*

$$\sum_{\substack{0 \leqslant h < q^\lambda \\ h \equiv a \bmod q^\delta}} \left\| F_{\mu,\lambda}(h) \right\|_{\mathbb{F}} \left\| F_{\lambda-\mu}(h+\ell) \right\|_{\mathbb{F}} \leqslant \Phi\left(q^\mu\right) \left\| F_\delta(a) \right\|_{\mathbb{F}} \left\| F_\delta(a+\ell) \right\|_{\mathbb{F}}. \tag{4.17}$$

Proof. If we write $h = uq^{\lambda-\mu} + v$, where $0 \leqslant u < q^\mu$ and $0 \leqslant v < q^{\lambda-\mu}$, then $h \equiv a \bmod q^\delta$ is equivalent to $v \equiv a \bmod q^\delta$ (we have $\delta \leqslant \lambda - \mu$). Denote the left-hand side of (4.17) by S. Then we obtain

$$\begin{aligned} S &= \sum_{\substack{0 \leqslant v < q^{\lambda-\mu} \\ v \equiv a \bmod q^\delta}} \sum_{0 \leqslant u < q^\mu} \left\| F_{\mu,\lambda}(uq^{\lambda-\mu} + v) \right\|_{\mathbb{F}} \left\| F_{\lambda-\mu}(uq^{\lambda-\mu} + v + \ell) \right\|_{\mathbb{F}} \\ &= \sum_{\substack{0 \leqslant v < q^{\lambda-\mu} \\ v \equiv a \bmod q^\delta}} \left\| F_{\lambda-\mu}(v) \right\|_{\mathbb{F}} \left\| F_{\lambda-\mu}(v+\ell) \right\|_{\mathbb{F}} q^{-\mu} \sum_{0 \leqslant u < q^\mu} \varphi_{q^\mu}\left(\frac{uq^{\lambda-\mu} + v}{q^\lambda} \right) \\ &\leqslant \sum_{\substack{0 \leqslant v < q^{\lambda-\mu} \\ v \equiv a \bmod q^\delta}} \left\| F_{\lambda-\mu}(v) \right\|_{\mathbb{F}} \left\| F_{\lambda-\mu}(v+\ell) \right\|_{\mathbb{F}} \Phi\left(q^\mu\right). \end{aligned}$$

Applying the Cauchy-Schwarz inequality yields

$$S \leqslant \Phi\left(q^\mu\right) \left(\sum_{\substack{0 \leqslant v < q^{\lambda-\mu} \\ v \equiv a \bmod q^\delta}} \left\| F_{\lambda-\mu}(v) \right\|_{\mathbb{F}}^2 \right)^{1/2} \cdot \left(\sum_{\substack{0 \leqslant v < q^{\lambda-\mu} \\ v \equiv a \bmod q^\delta}} \left\| F_{\lambda-\mu}(v+\ell) \right\|_{\mathbb{F}}^2 \right)^{1/2},$$

and Lemma 4.18 implies the desired result. □

Lemma 4.23. *Let $q \geqslant 2$, $a \in \mathbb{Z}$, $1 \leqslant \mu < \lambda$ and $\lambda - \mu \leqslant \delta \leqslant \lambda$. Then we have*

$$\sum_{\substack{0 \leqslant h_1, h_2 < q^\lambda \\ h_1 + h_2 \equiv a \bmod q^\delta}} \left\| F_{\mu,\lambda}(h_1) \right\|_{\mathbb{F}} \left\| F_{\mu,\lambda}(-h_2) \right\|_{\mathbb{F}} \leqslant d\, \Phi(q^{\lambda-\delta})\, \Phi(q^\mu). \tag{4.18}$$

Proof. We have

$$\begin{aligned} &\sum_{\substack{0 \leqslant h_1, h_2 < q^\lambda \\ h_1 + h_2 \equiv a \bmod q^\delta}} \left\| F_{\mu,\lambda}(h_1) \right\|_{\mathbb{F}} \left\| F_{\mu,\lambda}(-h_2) \right\|_{\mathbb{F}} \\ &\qquad = \sum_{0 \leqslant h_2 < q^\lambda} \left\| F_{\mu,\lambda}(-h_2) \right\|_{\mathbb{F}} \sum_{\substack{0 \leqslant h_1 < q^\lambda \\ h_1 \equiv -h_2 + a \bmod q^\delta}} \left\| F_{\mu,\lambda}(h_1) \right\|_{\mathbb{F}} \\ &\qquad \leqslant \Phi(q^{\lambda-\delta}) \sum_{0 \leqslant h_2 < q^\lambda} \left\| F_{\mu,\lambda}(-h_2) \right\|_{\mathbb{F}} \left\| F_{\lambda-\mu}(-h_2 + a) \right\|_{\mathbb{F}}. \end{aligned}$$

To obtain the last inequality, we employed Lemma 4.21 (note, that $\varphi_{q^{\mu-\lambda+\delta}}(\,\cdot\,) \leqslant q^{\mu-\lambda+\delta}$). Since we have

$$\left\| F_0(u) \right\|_{\mathbb{F}} = \left\| I_d \right\|_{\mathbb{F}} = \sqrt{d}$$

for all $u \in \mathbb{Z}$, Lemma 4.22 (with $\delta = 0$) yields the desired result. □

4.4 Proof of Theorem 4.1

In the next two sections we show that

$$\frac{1}{N}\sum_{0\leqslant n<N} D(T(n^2))$$

converges for every irreducible and unitary representation D (as N goes to infinity). With the help of this result, we will finish the proof of Theorem 4.1 in Section 4.4.3.

4.4.1 Irreducible representations of the form D_k

In this section we consider the m representations $D_0,\ldots,D_{m-1}$. Recall that these satisfy $D_k(T(v)) = \mathrm{e}(-kv/m)$ for all $v \geqslant 0$. We have

$$\begin{aligned}\frac{1}{N}\sum_{n=0}^{N-1} D_k(T(n^2)) &= \frac{1}{N}\sum_{n=0}^{N-1} \mathrm{e}\left(-\frac{kn^2}{m}\right)\\ &= \frac{1}{N}\left\lfloor\frac{N}{m}\right\rfloor\sum_{n=0}^{m-1}\mathrm{e}\left(-\frac{kn^2}{m}\right) + \frac{1}{N}\sum_{n=m\lfloor\frac{N}{m}\rfloor}^{N-1}\mathrm{e}\left(-\frac{kn^2}{m}\right).\end{aligned}$$

It follows that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} D_k(T(n^2)) = \frac{1}{m}G(-k,m), \tag{4.19}$$

where $G(a,c) = G(a,0;c)$ and $G(a,b;c)$ denotes the quadratic Gauss sum

$$G(a,b;c) = \sum_{n=0}^{c-1}\mathrm{e}\left(\frac{an^2+bn}{c}\right), \tag{4.20}$$

see Appendix A.3.

4.4.2 Irreducible representations different from D_k

Let D be an irreducible and unitary representation of degree $d \geqslant 1$ such that $D \neq D_k$ for all $0 \leqslant k < m$. Furthermore, let $\nu \in \mathbb{Z}^+$ be defined by $q^{\nu-1} < N \leqslant q^{\nu}$ and set

$$S_1 = \sum_{0\leqslant n<N} D(T(n^2)). \tag{4.21}$$

Then we will show that there exists a constant $\sigma > 0$, such that $\|S_1\|_{\mathbb{F}} \ll q^{(1-\sigma)\nu}$. In particular, this implies that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} D(T(n^2)) = 0.$$

We begin with applying Lemma 4.16 with $B = q^\nu$, $R = q^\rho$ and $Z(n) = D(T(n^2))$, where ρ is an integer satisfying $1 \leqslant \rho \leqslant \nu/2$ (note, that $\|D(T(n^2))\|_{\mathbb{F}} = \sqrt{d}$ for all $n \geqslant 0$). We can write

$$\begin{aligned}\|S_1\|_{\mathbb{F}} &\ll \left(\frac{N}{q^\rho} \sum_{|r|<q^\rho} \left(1 - \frac{|r|}{q^\rho}\right) \left\| \sum_{0 \leqslant n, n+r \leqslant q^\nu} D(T((n+r)^2)) D(T(n^2))^H \right\|_{\mathbb{F}} \right)^{1/2} + q^\rho \\ &\ll \left(q^{2\nu-\rho} + q^\nu \max_{1 \leqslant |r| < q^\rho} \left\| \sum_{0 \leqslant n, n+r \leqslant q^\nu} D(T((n+r)^2)) D(T(n^2))^H \right\|_{\mathbb{F}} \right)^{1/2} + q^\rho.\end{aligned}$$

In the last step, we separated the case $r = 0$ and $r \neq 0$. Additionally, we get an error term $O(q^{\nu+\rho})$ (inside the square root) when removing the summation condition $0 \leqslant n + r \leqslant q^\nu$. Since we have assumed $\rho \leqslant \nu/2$, this term can be neglected. Hence, we obtain

$$\|S_1\|_{\mathbb{F}} \ll q^{\nu-\rho/2} + q^{\nu/2} \max_{1 \leqslant |r| < q^\rho} \left\| \sum_{0 \leqslant n \leqslant q^\nu} D(T((n+r)^2)) D(T(n^2))^H \right\|_{\mathbb{F}}^{1/2}.$$

We set

$$\lambda := \nu + 2\rho + 1.$$

Recall that we have (using the fact that $g_0 = e$)

$$T_\lambda(n) := g_{\varepsilon_0(n)} g_{\varepsilon_1(n)} \cdots g_{\varepsilon_{\lambda-1}(n)},$$

where $\varepsilon_0(n), \ldots, \varepsilon_{\lambda-1}(n)$ are the λ lower placed digits in the q-ary digital expansion (with possibly leading zeros) of n.

Lemma 4.24. *For all integers ν and ρ with $\nu \geqslant 2$ and $1 \leqslant \rho \leqslant \nu/2$ and for all $r \in \mathbb{Z}$ with $|r| < q^\rho$, we denote by $E(r, \nu, \rho)$ the number of integers n such that $0 \leqslant n \leqslant q^\nu$ and*

$$D(T_\lambda((n+r)^2)) D(T_\lambda(n^2))^H \neq D(T((n+r)^2)) D(T(n^2))^H.$$

Then we have

$$E(r, \nu, \rho) \ll q^{\nu-\rho}.$$

Proof. In order to prove this lemma we recall what Mauduit and Rivat actually proved in [MR09, Lemme 16]. Let us fix an integer r satisfying $|r| < q^\rho$ and denote by $F(r, \nu, \rho)$ the number of integers $0 \leqslant n \leqslant q^\nu$ such that not all digits of n^2 and $(n+r)^2$ which are higher placed than $\lambda - 1$ are equal. Then it follows from their reasoning that $F(r, \nu, \rho) \ll q^{\nu-\rho}$. (Note, that the additional condition $q^{\nu-1} < n$ in their statement is not needed.) Using this fact, we can complete our proof. Let

$n^2 = \varepsilon_{\ell-1}\varepsilon_{\ell-2}\dots\varepsilon_\lambda\dots\varepsilon_0$, where $\ell > \lambda$ and with possibly leading zeros. If all digits of n^2 and $(n+r)^2$ which are higher placed than $\lambda - 1$ are equal, we have

$$\begin{aligned} &D(T((n+r)^2))D(T(n^2))^H \\ &\qquad = D(T_\lambda((n+r)^2))D(g_{\varepsilon_\lambda}\cdots g_{\varepsilon_{\ell-1}})D(g_{\varepsilon_\lambda}\cdots g_{\varepsilon_{\ell-1}})^H D(T_\lambda(n^2))^H \\ &\qquad = D(T_\lambda((n+r)^2))D(T_\lambda(n^2))^H. \end{aligned}$$

Here we used that D is a unitary representation. All matrices that come from the higher placed digits “cancel” out. We can bound $E(r,\nu,\rho)$ by $F(r,\nu,\rho)$ which proves the desired result. □

This lemma enables us to replace $D(T(u^2))$ by $D(T_\lambda(u^2))$. We obtain

$$\|S_1\|_{\mathbb{F}} \ll q^{\nu-\rho/2} + q^{\nu/2} \max_{1\leqslant |r| < q^\rho} \|S_2\|_{\mathbb{F}}^{1/2}, \tag{4.22}$$

where

$$S_2 := \sum_{0\leqslant n\leqslant q^\nu} D(T_\lambda((n+r)^2))D(T_\lambda(n^2))^H.$$

In view of Lemma 4.17, we set $Z(n) = D(T_\lambda((n+r)^2))D(T_\lambda(n^2))^H$, $N = q^\nu$, $S = q^{2\rho}$ and $k = q^\mu$, where μ is an integer satisfying

$$1 \leqslant \mu \leqslant \nu - 2\rho - 1.$$

Then we have to consider expressions of the form

$$\operatorname{tr}\left(D(T_\lambda((n+r+sq^\mu)^2))D(T_\lambda((n+sq^\mu)^2))^H D(T_\lambda(n^2))D(T_\lambda((n+r)^2))^H\right). \tag{4.23}$$

Using that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ for two matrices A and B, this is the same as

$$\operatorname{tr}\left(D(T_\lambda((n+r)^2))^H D(T_\lambda((n+r+sq^\mu)^2))D(T_\lambda((n+sq^\mu)^2))^H D(T_\lambda(n^2))\right).$$

Next, we recall that

$$T_{\mu,\lambda}(n) = g_{\varepsilon_\mu(n)}g_{\varepsilon_{\mu+1}(n)}\cdots g_{\varepsilon_{\lambda-1}(n)}.$$

Since

$$(n+r+sq^\mu)^2 = (n+r)^2 + q^\mu(2s(n+r) + s^2q^\mu),$$

we see that $(n+r+sq^\mu)^2$ and $(n+r)^2$ have the same μ lower placed digits. Using that D is a unitary representation (compare also with the proof of Lemma 4.24), we get

$$D(T_\lambda((n+r)^2))^H D(T_\lambda((n+r+sq^\mu)^2)) = D(T_{\mu,\lambda}((n+r)^2))^H D(T_{\mu,\lambda}((n+r+sq^\mu)^2)),$$

since the terms coming from the μ lower placed digits cancel out. The same argument works for $(n+sq^\mu)^2$ and n^2 and we have

$$D(T_\lambda((n+sq^\mu)^2))^H D(T_\lambda(n^2)) = D(T_{\mu,\lambda}((n+sq^\mu)^2))^H D(T_{\mu,\lambda}(n^2)).$$

Thus, we obtain that (4.23) can be written as

$$\operatorname{tr}\left(D(T_{\mu,\lambda}((n+r)^2))^H D(T_{\mu,\lambda}((n+r+sq^\mu)^2)) D(T_{\mu,\lambda}((n+sq^\mu)^2))^H D(T_{\mu,\lambda}(n^2))\right).$$

We set $I_{\nu,s,\mu} := \{0 \leqslant n \leqslant q^\nu : 0 \leqslant n + sq^\mu \leqslant q^\nu\}$. Then we finally have (cf. (4.11))

$$\begin{aligned}
&\left|\sum_{n\in I_{\nu,s,\mu}} \operatorname{tr}\left(D(T_\lambda((n+r)^2))^H D(T_\lambda((n+r+sq^\mu)^2)) D(T_\lambda((n+sq^\mu)^2))^H D(T_\lambda(n^2))\right)\right| \\
&\quad = \left|\operatorname{tr}\left(\sum_{n\in I_{\nu,s,\mu}}\left(D(T_{\mu,\lambda}((n+r)^2))^H D(T_{\mu,\lambda}((n+r+sq^\mu)^2))\right.\right.\right. \\
&\qquad\qquad\qquad\qquad \left.\left.\left. \cdot D(T_{\mu,\lambda}((n+sq^\mu)^2))^H D(T_{\mu,\lambda}(n^2))\right)\right)\right| \\
&\quad \leqslant \sqrt{d}\left\|\sum_{n\in I_{\nu,s,\mu}} D(T_{\mu,\lambda}((n+r)^2))^H D(T_{\mu,\lambda}((n+r+sq^\mu)^2))\right. \\
&\qquad\qquad\qquad\qquad \left. \cdot D(T_{\mu,\lambda}((n+sq^\mu)^2))^H D(T_{\mu,\lambda}(n^2))\right\|_{\mathbb{F}}.
\end{aligned}$$

Hence Lemma 4.17 gives

$$\begin{aligned}
\|S_2\|_{\mathbb{F}}^2 &\ll q^{\nu-2\rho} \sum_{|s|<q^{2\rho}} \left(1 - \frac{|s|}{q^{2\rho}}\right) \|S_3\|_{\mathbb{F}} \\
&\ll q^{2\nu-2\rho} + q^\nu \max_{1\leqslant|s|<q^{2\rho}} \|S_3\|_{\mathbb{F}}, \qquad (4.24)
\end{aligned}$$

where S_3 denotes the sum

$$\sum_{n\in I_{\nu,s,\mu}} D(T_{\mu,\lambda}((n+r)^2))^H D(T_{\mu,\lambda}((n+r+sq^\mu)^2)) D(T_{\mu,\lambda}((n+sq^\mu)^2))^H D(T_{\mu,\lambda}(n^2)).$$

The inverse of the Fourier term $F_{\mu,\lambda}$ (defined in Section 4.3.2) is given by

$$D(T_{\mu,\lambda}(u)) = \sum_{0\leqslant h<q^\lambda} F_{\mu,\lambda}(h)\,\mathrm{e}\left(\frac{uh}{q^\lambda}\right).$$

Hence we obtain

$$\begin{aligned}
S_3 = &\sum_{0\leqslant h_1,h_2,h_3,h_4<q^\lambda} F_{\mu,\lambda}(-h_1)^H F_{\mu,\lambda}(h_2) F_{\mu,\lambda}(-h_3)^H F_{\mu,\lambda}(h_4) \\
&\cdot \sum_{n\in I_{\nu,s,\mu}} \mathrm{e}\left(\frac{h_1(n+r)^2 + h_2(n+r+sq^\mu)^2 + h_3(n+sq^\mu)^2 + h_4 n^2}{q^\lambda}\right). \qquad (4.25)
\end{aligned}$$

Assume that $c \geqslant 2$ is an integer and $(z_n)_{n\in\mathbb{Z}}$ is a sequence of complex numbers that is periodic of period c. It is shown in [MR09, Lemme 18], that one has for all M_1, $M_2 \in \mathbb{Z}$ with $1 \leqslant M_2 \leqslant c$ the estimate

$$\left|\sum_{M_1<n\leqslant M_1+M_2} z_n\right| \leqslant \frac{2}{\pi}\log\left(\frac{4e^{\pi/2}c}{\pi}\right)\max_{0\leqslant \ell<c}\left|\sum_{0\leqslant n<c} z_n \,\mathrm{e}\left(\frac{\ell n}{c}\right)\right|.$$

Next we apply this to (4.25) with $c = q^\lambda$ and M_1 and M_2 chosen appropriately. We then get

$$\begin{aligned}\|S_3\|_{\mathbb{F}} \leqslant \frac{2}{\pi}\log\left(\frac{4e^{\pi/2}q^\lambda}{\pi}\right)&\max_{0\leqslant \ell<q^\lambda}\sum_{d\mid q^\lambda}\\ &\cdot\sum_{\substack{0\leqslant h_1,h_2,h_3,h_4<q^\lambda\\(h_1+h_2+h_3+h_4,q^\lambda)=d}}\|F_{\mu,\lambda}(h_1)\|_{\mathbb{F}}\,\|F_{\mu,\lambda}(h_2)\|_{\mathbb{F}}\,\|F_{\mu,\lambda}(h_3)\|_{\mathbb{F}}\,\|F_{\mu,\lambda}(h_4)\|_{\mathbb{F}}\\ &\cdot|G(h_1+h_2+h_3+h_4,2r(h_1+h_2)+2sq^\mu(h_2+h_3)+l;q^\lambda)|,\end{aligned}$$

where $G(a,b;c)$ denotes the quadratic Gauss sum defined in (4.20). Proposition A.11 implies

$$\begin{aligned}\|S_3\|_{\mathbb{F}} \leqslant \frac{2\sqrt{2}}{\pi}\log\left(\frac{4e^{\pi/2}q^\lambda}{\pi}\right)&q^{\lambda/2}\max_{0\leqslant \ell<q^\lambda}\sum_{d\mid q^\lambda}d^{1/2}\\ &\cdot\sum_{\substack{0\leqslant h_1,h_2,h_3,h_4<q^\lambda\\(h_1+h_2+h_3+h_4,q^\lambda)=d\\d\mid 2r(h_1+h_2)+2sq^\mu(h_2+h_3)+\ell}}\|F_{\mu,\lambda}(h_1)\|_{\mathbb{F}}\,\|F_{\mu,\lambda}(h_2)\|_{\mathbb{F}}\,\|F_{\mu,\lambda}(h_3)\|_{\mathbb{F}}\,\|F_{\mu,\lambda}(h_4)\|_{\mathbb{F}}\,.\end{aligned}\tag{4.26}$$

The same way as in the work of Mauduit and Rivat [MR09, Section 5.5 - 5.8], one can now show that the integer μ can be chosen in such a way that

$$\|S_3\|_{\mathbb{F}} \ll \nu^{\omega(q)+6}q^{\nu-2\rho},\tag{4.27}$$

whenever ρ is an integer smaller than ν times some constant only depending on the chosen representation D. (The constant $\omega(q)$ denotes the number of distinct prime divisors of q). In the following, we give a short overview of the reasoning how the estimates of the Fourier transform proved in Section 4.3.2 imply (4.27). Compare also with Chapter 3 (especially Section 3.4.3), where the corresponding estimates are simpler than the one given below, since we only used once a van der Corput-type inequality. Since the method of the proof of (4.27) is exactly the same as in [MR09], we omit most of the details and just show where the Fourier transform estimates come into play.

First, we replace the condition $(h_1+h_2+h_3+h_4,q^\lambda) = d$ by the less restrictive one $h_1+h_2+h_3+h_4 \equiv 0 \bmod d$. Next, we consider the sum over d separately for

$\nu_q(d) < \Delta$, $\Delta \leqslant \nu_q(d) < \mu$ and $\mu \leqslant \nu_q(d) \leqslant \lambda$, where Δ is an integer satisfying $1 \leqslant \Delta < \mu$. We write

$$\|S_3\|_{\mathbb{F}} \ll_q \lambda q^{\lambda/2} \max_{0\leqslant \ell < q^\lambda} (S_4 + S_5 + S_6),$$

where S_4 is equal to

$$\sum_{\substack{d|q^\lambda \\ \nu_q(d)<\Delta}} d^{1/2} \sum_{\substack{0\leqslant h_1,h_2,h_3,h_4<q^\lambda \\ h_1+h_2+h_3+h_4\equiv 0 \bmod d \\ 2r(h_1+h_2)+2sq^\mu(h_2+h_3)+\ell\equiv 0 \bmod d}} \|F_{\mu,\lambda}(h_1)\|_{\mathbb{F}} \|F_{\mu,\lambda}(h_2)\|_{\mathbb{F}} \|F_{\mu,\lambda}(h_3)\|_{\mathbb{F}} \|F_{\mu,\lambda}(h_4)\|_{\mathbb{F}},$$

S_5 is equal to

$$\sum_{\substack{d|q^\lambda \\ \Delta\leqslant\nu_q(d)<\mu}} d^{1/2} \sum_{\substack{0\leqslant h_1,h_2,h_3,h_4<q^\lambda \\ h_1+h_2+h_3+h_4\equiv 0 \bmod d \\ 2r(h_1+h_2)+2sq^\mu(h_2+h_3)+\ell\equiv 0 \bmod d}} \|F_{\mu,\lambda}(h_1)\|_{\mathbb{F}} \|F_{\mu,\lambda}(h_2)\|_{\mathbb{F}} \|F_{\mu,\lambda}(h_3)\|_{\mathbb{F}} \|F_{\mu,\lambda}(h_4)\|_{\mathbb{F}},$$

and S_6 is equal to

$$\sum_{\substack{d|q^\lambda \\ \nu_q(d)\geqslant\mu}} d^{1/2} \sum_{\substack{0\leqslant h_1,h_2,h_3,h_4<q^\lambda \\ h_1+h_2+h_3+h_4\equiv 0 \bmod d \\ 2r(h_1+h_2)+2sq^\mu(h_2+h_3)+\ell\equiv 0 \bmod d}} \|F_{\mu,\lambda}(h_1)\|_{\mathbb{F}} \|F_{\mu,\lambda}(h_2)\|_{\mathbb{F}} \|F_{\mu,\lambda}(h_3)\|_{\mathbb{F}} \|F_{\mu,\lambda}(h_4)\|_{\mathbb{F}}.$$

In order to find an upper bound of S_4, we ignore the additional conditions for the indices $h_1, \ldots, h_4$. Lemma 4.20 and some simple calculations show that this sum is negligible if Δ is not too large (for the exact value of the chosen integer Δ see [MR09, Equation (51)]).

As to bound S_5, one has to rewrite the summation conditions and an application of Lemma 4.23 shows that this sum is negligible, too. Here we use that $\delta = \nu_q(d)$ is not too small, but also not too large.

The estimation of the sum S_6 is the most difficult part of the verification of (4.27). Set

$$\rho_q := \left\lfloor \rho \frac{\log q}{\log 2} + 1 \right\rfloor.$$

First one shows that there exist integers s'' and ℓ'', which depend on r, s, d and ℓ, such that (we use the notation $\delta = \nu_q(d)$)

$$S_6 \leqslant \sum_{\substack{d|q^\lambda \\ \nu_q(d)\geqslant\mu}} d^{1/2}(S_7 + S_8),$$

where

$$S_7 = \sum_{\substack{0\leqslant h_1,h_2,h_3,h_4<q^\lambda \\ h_1+h_2+s''q^{\mu-\rho_q}(h_1+h_3)+\ell''\equiv 0 \bmod q^{\delta-\rho_q} \\ h_3+h_4-s''q^{\mu-\rho_q}(h_1+h_3)-\ell''\equiv 0 \bmod q^{\delta-\rho_q} \\ \left\| \frac{h_1+s''q^{\eta-\rho_q}(h_1+h_3)+\ell''}{q^{\delta-\rho_q}} \right\| \geqslant q^{-\eta+\lambda-\delta+\rho_q+4\rho}}} \|F_{\mu,\lambda}(h_1)\|_{\mathbb{F}} \|F_{\mu,\lambda}(h_2)\|_{\mathbb{F}} \|F_{\mu,\lambda}(h_3)\|_{\mathbb{F}} \|F_{\mu,\lambda}(h_4)\|_{\mathbb{F}},$$

and

$$S_8 = \sum_{\substack{0 \leqslant h_1,h_2,h_3,h_4 < q^\lambda \\ h_1+h_2+s''q^{\mu-\rho_q}(h_1+h_3)+\ell'' \equiv 0 \bmod q^{\delta-\rho_q} \\ h_3+h_4-s''q^{\mu-\rho_q}(h_1+h_3)-\ell'' \equiv 0 \bmod q^{\delta-\rho_q} \\ \left\| \frac{h_1+s''q^{\eta-\rho_q}(h_1+h_3)+\ell''}{q^{\delta-\rho_q}} \right\| < q^{-\eta+\lambda-\delta+\rho_q+4\rho}}} \|F_{\mu,\lambda}(h_1)\|_{\mathbb{F}} \, \|F_{\mu,\lambda}(h_2)\|_{\mathbb{F}} \, \|F_{\mu,\lambda}(h_3)\|_{\mathbb{F}} \, \|F_{\mu,\lambda}(h_4)\|_{\mathbb{F}} \,.$$

The sum S_7 can be estimated by employing Lemma 4.21 twice and Lemma 4.22 once. The crucial part of this estimation is the fact that the summation conditions of the sum S_7 imply that

$$q^{-\mu+\lambda-\delta+\rho_q} \varphi_{q^{\mu-\lambda+\delta-\rho_q}} \left(\frac{h_1 + s''q^{\mu-\rho_q}(h_1+h_3)+\ell''}{q^{\delta-\rho_q}} \right) \leqslant \frac{q^{-4\rho}}{2},$$

which finally shows that the part of S_6 coming from S_7 is negligible. For the calculation of S_8 we set

$$\rho'_q := \left\lfloor 2\rho \frac{\log q}{\log 2} + 1 \right\rfloor .$$

If $\delta - \mu - \rho_q - \rho'_q < 0$, then we can employ again Lemma 4.21 twice and Lemma 4.22 once. However, this time these steps work since the summation interval of δ is very short and away from λ (if we sum up the part in S_6 coming from S_8 with the additional condition $\delta - \mu - \rho_q - \rho'_q < 0$).

If $\delta - \mu - \rho_q - \rho'_q \geqslant 0$, some nontrivial considerations using congruence relations also yield an negligible upper bound for the rest of the sum S_8 (see [MR09, Section 5.7]). One has to employ Lemma 4.21 twice with different parameters in the stated lemma and Lemma 4.22 once. Furthermore, one has to use the L^∞-norm estimate of $\|F_\lambda(\,\cdot\,)\|$ proved in Section 4.2 (see Lemma 4.13).

Combining (4.22), (4.24) and (4.27), we are done since we can choose ρ such that

$$\|S_1\|_{\mathbb{F}} \ll q^{(1-\sigma)\nu}$$

for some constant $\sigma > 0$.

Finally, we want to remark that the L^∞-norm estimate is only used in the last step of the proof. Note, that except for this step the presented method works also for the representations D_k, $0 \leqslant k < m$ (the L^∞-norm estimate does not hold true for them, see Remark 4.14). Interestingly, it is proved in Section 4.4.1 that for these representations the statement of this section is false in general.

4.4.3 Final steps in the proof of Theorem 4.1

Let m be the characteristic integer of $g_0, \ldots, g_{q-1}$ and q. Lemma 4.12 implies that the set $U := \mathrm{cl}(\{T(mn) : n \geqslant 0\})$ is a normal subgroup of G (of index m) with cosets $g_u U = \mathrm{cl}(\{T(mn+u) : n \geqslant 0\})$, $0 \leqslant u < m$. Let us define the function f by

$$f(g) = 1 + \frac{1}{m} G(1,m) D_1(g) + \cdots + \frac{1}{m} G(m-1,m) D_{m-1}(g).$$

Since the representations are continuous, f is a continuous function. Moreover, it is a positive function and it satisfies

$$f(v) = \sum_{u=0}^{m-1} \mathbf{1}_{g_u U}(v) \cdot Q(u,m). \tag{4.28}$$

In order to see this, we state and prove the following lemma:

Lemma 4.25. *Let $c \geqslant 1$, $a \in \mathbb{Z}$ and set $Q(a,c) = \#\{0 \leqslant n < c : n^2 \equiv a \bmod c\}$. Then we have*

$$Q(a,c) = \frac{1}{c} \sum_{k=0}^{c-1} G(k,c)\, \mathrm{e}\left(-\frac{ka}{c}\right).$$

Proof. Let us consider the group $\mathbb{Z}/c\mathbb{Z}$ (with $+$ as the group operation). The representations of $\mathbb{Z}/c\mathbb{Z}$ are given by

$$\chi_k(u) = \mathrm{e}\left(\frac{ku}{c}\right)$$

for all $u \in \mathbb{Z}/c\mathbb{Z}$ and all $0 \leqslant k < c$. For any $n \in \mathbb{N}$ let x_n be the element of $\mathbb{Z}/c\mathbb{Z}$ defined by $n \equiv x_n \bmod c$. Then we obtain

$$\lim_{N\to\infty} \frac{1}{N} \#\{0 \leqslant n < N : x_{n^2} \equiv a \bmod c\} = \frac{1}{c} Q(a,c).$$

Thus, the sequence $(x_{n^2})_{n \geqslant 0}$ is $\nu_{\mathbb{Z}/c\mathbb{Z}}$-uniformly distributed in $\mathbb{Z}/c\mathbb{Z}$, where the measure $\nu_{\mathbb{Z}/c\mathbb{Z}}$ is defined by

$$\nu_{\mathbb{Z}/c\mathbb{Z}}(v) = \frac{1}{c} Q(v,c).$$

It follows from Lemma 4.8 that for every $0 \leqslant k < c$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \chi_k(x_{n^2}) = \int_{\mathbb{Z}/c\mathbb{Z}} \chi_k \mathrm{d}\nu_{\mathbb{Z}/c\mathbb{Z}}. \tag{4.29}$$

As in Section 4.4.1, we obtain

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \chi_k(x_{n^2}) = \frac{1}{c} G(k,c).$$

On the other hand, we have

$$\int_{\mathbb{Z}/c\mathbb{Z}} \chi_k \mathrm{d}\nu_{\mathbb{Z}/c\mathbb{Z}} = \frac{1}{c} \sum_{v=0}^{c-1} \mathrm{e}\left(-\frac{vk}{c}\right) Q(v,c).$$

Summing up the left and right-hand side of (4.29) from $k=0$ to $k=c-1$ (weighted with $\mathrm{e}(-ka/c)$), we obtain

$$\frac{1}{c} \sum_{k=0}^{c-1} G(k,c)\, \mathrm{e}\left(-\frac{ka}{c}\right) = \frac{1}{c} \sum_{k=0}^{c-1} \mathrm{e}\left(-\frac{ka}{c}\right) \sum_{v=0}^{c-1} \mathrm{e}\left(\frac{vk}{c}\right) Q(v,c). \tag{4.30}$$

The right-hand side of (4.30) is equal to

$$\sum_{v=0}^{c-1} Q(v,c) \frac{1}{c} \sum_{k=0}^{c-1} \mathrm{e}\left(\frac{k(v-a)}{c}\right) = Q(a,c).$$

This shows the desired result. □

If n and v are integers $\geqslant 0$, then

$$f(T(nm+v)) = \frac{1}{m} \sum_{k=0}^{m-1} G(k,m) D_k(T(nm+v)) = \frac{1}{m} \sum_{k=0}^{m-1} G(k,m) \, \mathrm{e}\left(-\frac{vk}{m}\right).$$

Employing Lemma 4.25, we obtain that

$$f(T(nm+v)) = Q(v,m).$$

Since $(T(n))_{n \geqslant 0}$ is dense in G, Equation (4.28) holds true, indeed. This allows us to define the measure

$$\mathrm{d}\nu = f \mathrm{d}\mu.$$

We proceed as in the linear case and show that $(T(n^2))_{n \geqslant 0}$ is ν-uniformly distributed in G. Let $\{D^\alpha,\ \alpha \in \mathcal{A}\}$ be again a complete set of pairwise inequivalent irreducible unitary representations and set $e_{ij}^\alpha(g) = \sqrt{n_\alpha} d_{ij}^\alpha(g)$, where $D^\alpha = (d_{ij})_{1 \leqslant i,j \leqslant n_\alpha}$ (recall, that the set $\{e_{ij}^\alpha\}$ forms a complete orthonormal system in the Hilbert space $L^2(G)$). We obtain for D_k, $k = 0, \ldots, m-1$ that

$$\int_G D_k f \mathrm{d}\mu = \sum_{\ell=0}^{m-1} \frac{1}{m} G(-\ell, m) \left\langle D_k, \overline{D_\ell} \right\rangle = \frac{1}{m} G(-k,m).$$

For all other representations $D^\alpha = (d_{ij}^\alpha)_{1 \leqslant i,j \leqslant n_\alpha}$, we get

$$\int_G d_{ij}^\alpha f \mathrm{d}\mu = \sum_{\ell=0}^{m-1} \frac{1}{m} G(-\ell, m) \left\langle d_{ij}^\alpha, \overline{D_\ell} \right\rangle = 0.$$

Finally, this proves Theorem 4.1 and completes the treatment of the generalized Thue-Morse sequence.

Chapter 5

Density properties of automatic sequences

An old French mathematician said:
"A mathematical theory is not to be considered complete
until you have made it so clear that you can explain it
to the first man whom you meet on the street."

David Hilbert (1862–1943)

Let $(u_n)_{n\in\mathbb{N}}$ be a q-automatic sequence with letters in an alphabet Δ. It is well-known that for every letter c in $(u_n)_{n\in\mathbb{N}}$ the logarithmic frequency

$$\lim_{x\to\infty} \frac{1}{\log x} \sum_{\substack{0\leqslant n<x\\ u_n=c}} \frac{1}{n}$$

exists. However, the ordinary frequency

$$\lim_{x\to\infty} \frac{1}{x}\#\{0\leqslant n<x : u_n=c\}$$

does not exist in general. (For the definition of q-automatic sequences and these facts see Section 5.1). What can we say about the frequency of a letter in a subsequence of $(u_n)_{n\in\mathbb{N}}$? Since linear subsequences $(u_{an+b})_{n\in\mathbb{N}}$ of automatic sequences are again automatic (see [AS03, Theorem 6.8.1]), the logarithmic frequency of a letter in $(u_{an+b})_{n\in\mathbb{N}}$ exists. Polynomial subsequences of higher order need not to be automatic any more. This makes the study of density properties in such subsequences much more difficult. Using the machinery of the generalized Thue-Morse sequence, we show that for invertible q-automatic sequences the frequency of each letter of the subsequence $(u_{n^2})_{n\in\mathbb{N}}$ exist. Moreover, we establish a connection between the frequency of letters in the q-automatic sequence $(u_n)_{n\in\mathbb{N}}$ and its subsequence $(u_{\lfloor n^c\rfloor})_{n\in\mathbb{N}}$ for $1<c<7/5$. For example, we show that the frequency of a letter a in $(u_n)_{n\in\mathbb{N}}$ exists if and only if the frequency of a exists in $(u_{\lfloor n^c\rfloor})_{n\in\mathbb{N}}$.

5.1 Some facts concerning automatic sequences

Automatic sequences $(u_n)_{n\geqslant 0}$ can be seen as the output sequence (or the image of the output sequence in an alphabet Δ) of a finite automaton when the input is the q-ary digital expansion of n (see for example [AS03, Lia90]). More precisely, a finite automaton has finitely many states. One starts in an initial state and then moves around the states depending on the input sequence (the q-ary digits of n). The moves are deterministic and can be encoded with the help of so-called *transition matrices* M_k, $k = 0, \dots, q-1$, which have the property that the entry m_{ij}^k (the i-th row and j-th column of the matrix M_k) is 1 if there is a move from state j to state i when the input digit equals k. All other entries are zero. (Note that the dimension of the matrices M_k is equal to the number of states.) Such a finite automaton can be represented by a directed graph, where a directed edge is labeled with a number k (between 0 and $q-1$) which indicates the new state if the input digit equals k. For every n, the automaton terminates at some state $s(n)$. Automatic sequences are now computed with the help of an output function that is defined on the states: $s(n) \mapsto u_n$. In Figure 5.1 we see the automaton that creates the Thue-Morse sequence. The two states of the automaton are s_1 (initial state) and s_2, the output function maps s_1 to 0 and s_2 to 1 and the transition matrices are given by

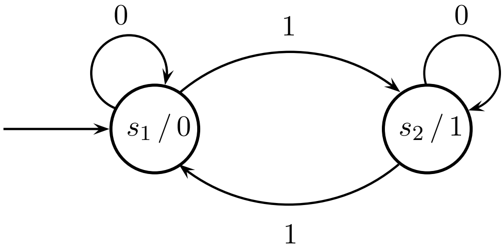


Figure 5.1: Automaton of the Thue-Morse sequence

$$M_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad M_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The automaton defined in Figure 5.2 has three states and the output function maps s_1 and s_2 to the letter a and s_3 to b. The transition matrices are given by

$$M_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad M_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \text{and} \qquad M_2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

The sequence starts with $aaaaabaabaabaaabba\dots$, and, as we will see later, it is related to the symmetric group S_3. There are several equivalent ways to describe automatic sequences (see [AS03], in particular Definition 5.1.1 and Theorem 5.2.1). For example, one can also describe q-automatic sequences in terms of morphisms

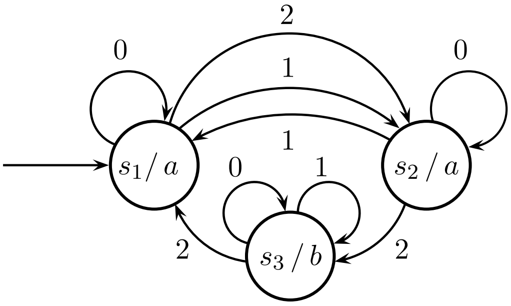


Figure 5.2: Automaton related to the S_3

(substitutions). Alternatively, by a theorem of Cobham a sequence $(u_n)_{n\geqslant 0}$ is q-automatic if and only if it is the image, under a coding, of a fixed point of a q-uniform morphism (compare with [AS03, Theorem 6.3.2]).

However, we will use the approach as described above. Let $\mathcal{A}_d$ be the set of all $d\times d$ matrices with the property that in each column there is exactly one entry equal to 1 and all other entries are 0. If the automaton has d states, then $M_k \in \mathcal{A}_d$ for every $0 \leqslant k < q$. Set $S(0) = M_0$ and

$$S(n) = M_{\varepsilon_0(n)} M_{\varepsilon_1(n)} \cdots M_{\varepsilon_{\ell-1}(n)} \tag{5.1}$$

for $n \geqslant 1$, where $(\varepsilon_{\ell-1}(n)\varepsilon_{\ell-2}(n)\dots\varepsilon_1(n)\varepsilon_0(n))_q$ denotes the q-ary digital expansion of n. Then we have that the last state reached is s_j (if the input is n) if and only if

$$S(n)e_1 = e_j.$$

Thus, a sequence $(u_n)_{n\geqslant 0}$ is a q-automatic sequence if and only if there exists q matrices in $\mathcal{A}_d$ (for some $d \geqslant 2$), such that u_n is given by the image of an output function (acting on $e_1,\dots,e_d$) of $S(n)e_1$. Additionally, if a letter a occurs in the automatic sequence $(u_n)_{n\geqslant 0}$, then there exists a vector $z = z_a \in \mathbb{C}^d$ which entries 0 and 1 such that

$$z^T S(n)e_1 = \begin{cases} 1, & \text{if } u_n = a, \\ 0, & \text{otherwise.} \end{cases} \tag{5.2}$$

Indeed, if a occurs in $(u_n)_{n\geqslant 0}$, then there exist some states $s_{\ell_1},\dots,s_{\ell_k}$ such that the output function maps these states to a. The vector $z = (z_1,\dots,z_d)^T$ with

$$z_i = \begin{cases} 1, & \text{if } i \in \{\ell_1,\dots,\ell_k\}, \\ 0, & \text{otherwise,} \end{cases}$$

does the job. At the end of this section we shortly recall the definition of the logarithmic and natural frequencies of letters. Let Δ be an alphabet and $(u_n)_{n\in\mathbb{N}}$ a sequence in Δ. Furthermore, let $a \in \Delta$. If the limit

$$\lim_{N\to\infty} \frac{1}{N} \#\{0 \leqslant n < N : u_n = a\}$$

exists, then it is called the frequency of a in $(u_n)_{n\in\mathbb{N}}$. A weaker version of density is the following one: If the limit

$$\lim_{x\to\infty}\frac{1}{\log x}\sum_{\substack{0\leqslant n<x\\ u_n=a}}\frac{1}{n}$$

exists, then it is called the logarithmic frequency of a in $(u_n)_{n\geqslant 0}$. If the frequency of a exists, then it is the same as the logarithmic frequency. As already mentioned, for arbitrary automatic sequences $(u_n)_{n\in\mathbb{N}}$ there need not exist the frequency of each letter. Take for example

$$M_0=\begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix}\qquad\text{and}\qquad M_1=\begin{pmatrix}0&0&0\\1&0&1\\0&1&0\end{pmatrix},$$

(cf. [AS03, Example 8.1.2]). Nevertheless, if $(u_n)_{n\in\mathbb{N}}$ is primitive (that is, the corresponding graph is strongly connected), then it is known that the frequencies of all letters exist. Moreover, one can calculated them effectively (see [AS03, Theorem 8.4.7]). Cobham showed, that the logarithmic frequency of each letter in an automatic sequence exists and Peter [Pet03] gave a necessary and sufficient condition for the existence of the ordinary frequency. Using our notation, he showed that the frequency of a letter a equals d, if and only if

$$z_a\left(\lim_{k\to\infty}\left(\frac{M_0+\ldots+M_{q-1}}{q}\right)^{mk}\right)e_i=d$$

for every $1\leqslant i\leqslant d$, where z_a is the vector defined in (5.2) and m is a positive integer such that the limit exists (note, that such an integer m always exists).

Furthermore, it is known that a subsequence of an automatic sequence of the form $(u_{an+b})_{n\geqslant 0}$ is again an automatic sequence (see [AS03, Theorem 6.8.1]). If we consider the subsequence $(u_{n^2})_{n\in\mathbb{N}}$, then it is of a completely different nature. If one takes for example the Thue-Morse sequence $(t_n)_{n\geqslant 1}$, then it follows by a Theorem of Allouche and Salon [AS93] that $(t_{n^2})_{n\in\mathbb{N}}$ is not 2-automatic. Moshe [Mos07] showed in a recent work the stronger result that the subword complexity of $(t_{n^2})_{n\in\mathbb{N}}$ is maximal, that is, every finite word appears as a subword. (Note, that the subword complexity of an automatic sequence is $O(n)$, see [AS03, Theorem 10.3.1]).

5.2 Automatic sequences and squares

The main result of this section is an application of Theorem 4.1 to so-called invertible automatic sequences (see Theorem 5.2). It turns out that there is a close relation between sequences $(T(n))_{n\in\mathbb{N}}$ of the form (4.1) and automatic sequences. In particular, generalized Thue-Morse sequences in the special case of finite groups are linked to matrices defined in (5.1).

Let G be a finite group of order $|G|$. Then G is also a compact topological group with respect to the discrete topology on G (every element is an open set). The Haar measure on G is the (normed) counting measure, that is

$$\mu(B) = \frac{1}{|G|}\,\#\{g : g \in B\}$$

for every $B \subseteq G$. Since one-dimensional representations of a finite group have to be $|G|$-th roots of unity, we see that the characteristic integer m has to divide $(|G|, q-1)$. If we take for example $G = \mathbb{Z}/r\mathbb{Z}$ and $g_j = j \bmod r$, $0 \leqslant j < q$, then we have $m = (r, q-1)$ and $T(n) = s_q(n) \bmod r$, where s_q denotes the q-ary sum-of-digits function. As already mentioned, this is exactly the case considered by Mauduit and Rivat in [MR09] (see Remark 4.3).

It is convenient to work with permutation matrices instead of abstract group elements. If G is a finite group, then G is isomorphic to a subgroup of the symmetric group $S_{|G|}$ (Cayley's Theorem). Thus, we can assume that $g_0, \dots, g_{q-1}$ are permutations in S_d for some integer $d \geqslant 1$ and G is a subgroup of S_d. The group G has a natural d-dimensional representation χ, the so-called *permutation representation.* It is defined as follows: Let $\pi \in S_d$, then

$$\chi(\pi) = \left(e_{\pi(1)}, \dots, e_{\pi(d)}\right),$$

where e_j denotes the j-th standard vector in $\mathbb{Z}^d$ (that is, all entries are 0 except the j-th, which is equal to 1). By definition it is clear that $\chi(\pi)$ is a permutation matrix, that is, $(x_1, \dots, x_d)\,\chi(\pi) = (x_{\pi(1)}, \dots, x_{\pi(d)})$. Obviously, permutation matrices are orthogonal (and unitary) matrices.

Definition 5.1. Let $(u_n)_{n\in\mathbb{N}}$ be a q-automatic sequence. Then we call $(u_n)_{n\in\mathbb{N}}$ an *invertible q-automatic sequence* if there exists an automaton such that all transition matrices are invertible and such that the transition matrix of zero is given by the identity matrix.

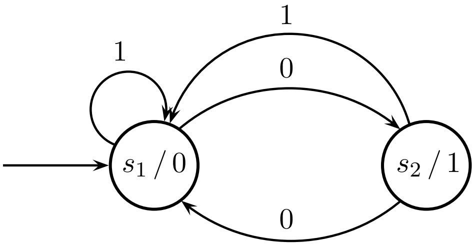


Figure 5.3: 2-adic valuation mod2

The set $\mathcal{A}_d$ is a monoid with respect to the matrix multiplication and all invertible matrices form a group (which is isomorphic to S_d). Taking this group as our group H, the matrices $M_0, \dots, M_{q-1}$ generate a subgroup G and we can use Theorem 4.1 to analyze the subsequence $(u_{n^2})_{n\in\mathbb{N}}$ of such invertible automatic sequences. As already indicated above, these matrices can be also seen as the permutation representation of the symmetric group S_d. Note, that the Thue-Morse sequence is an invertible 2-automatic sequence. The sequence induced by the automaton given in Figure 5.2 is an invertible 3-automatic sequence. (The transition matrices can be seen as the permutation representation of the identity element, the 2-cycle (12) and the 3-cycle (123) in S_3). If the matrices are interpreted as elements of the S_3, then they generate the whole group (that is, $G = H$ is isomorphic to S_3).

The following Theorem is a direct consequence of Theorem 4.1.

Theorem 5.2. *Let $q \geqslant 2$ and $(u_n)_{n\in\mathbb{N}}$ be an invertible q-automatic sequence. Then the frequency of each letter of the subsequence $(u_{n^2})_{n\in\mathbb{N}}$ exists.*

Remark 5.3. If the output function and the corresponding group generated by the transition matrices are known, then the exact frequencies of all letters can be given. Furthermore, one can show that if the output function is trivial (i.e., every state is mapped to a different letter in Δ) and the graph is strongly connected, then the frequencies are all equal if the characteristic number of the underlying group is $\leqslant 2$.

Remark 5.4. It follows from the last remark that if $(u_n)_{n\in\mathbb{N}}$ is a primitive 2-automatic or 3-automatic invertible sequence, then all the frequencies exist and they are strictly positive (since in this case $m \leqslant 2$). Actually, it is not hard to see that the frequencies are the same as for the original sequence. If the automatic sequence is not invertible this needs not to be true anymore: For instance, consider the sequence $(a_n)_{n\in\mathbb{N}}$ defined by the automaton given in Figure 5.3. We have $a_0 = 1$ and a_n is equal to $\nu_2(n) \bmod 2$ for $n \geqslant 1$, where $\nu_2(n)$ denotes the 2-adic valuation of n. Although the frequencies of 0 and 1 are strictly positive (2/3 and 1/3), we have $a_{n^2} = 0$ for all $n \geqslant 1$.

Remark 5.5. Let $(u_n)_{n\in\mathbb{N}}$ be a invertible q-automatic sequence such that the corresponding graph is strongly connected. Suppose that this graph has d states. We then know that each state is an output state with frequency $1/d$ (see [AS03, Theorem 8.4.7]). On the other hand, a similar argumentation as in the proof of Theorem 5.2 (for the sequence $(u_n)_{n\in\mathbb{N}}$ instead of $(u_{n^2})_{n\in\mathbb{N}}$) shows that this frequency is also given by p/s, where p is some positive integer and s is the cardinality of the group generated by the transition matrices. Thus, it follows that d has to be a divisor of s. On the other hand, s is for sure a divisor of $d!$ since the group generated by the transition matrices is isomorph to a subgroup of the permutation group with d elements.

Remark 5.6. If we take for example the automatic sequence generated by the automaton given in Figure 5.2, then we obtain that the underlying group is the symmetric group S_3 and $m = 2$. Thus, taking into account the special output function, the subsequence generated by the squares has the property, that the letter a has frequency 2/3 and the letter b has frequency 1/3.

Remark 5.7. It would be interesting to show a similar result for a wider range of automatic sequences. For special sequences like the Rudin-Shapiro sequence generated by the automaton given in Figure 5.4 this can be done, but a more general result seems to be difficult to prove (one cannot use the group theoretic structure of invertible sequences as presented in the previous chapter).

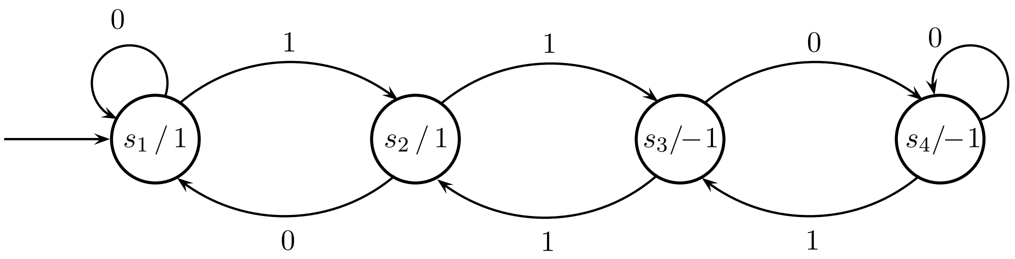


Figure 5.4: The Rudin-Shapiro sequence

Proof. We can assume that there exists an automaton with d states and with invertible transition matrices M_k for $k = 0, \ldots, k-1$ such that $M_0 = I_d$. The matrix sequence $(S(n))_{n\in\mathbb{N}}$ defined above coincides with the sequence $(T(n))_{n\in\mathbb{N}}$ defined in (4.1) (if we set $g_k = M_k$ for $0 \leqslant k < q$). Let the different elements in the subgroup generated by $M_0, \ldots, M_{q-1}$ be denoted by $N_0, \ldots, N_{s-1}$. Then

$$\sum_{n=0}^{N-1} T(n^2) = \sum_{r=0}^{s-1} \#\{0 \leqslant n < N : T(n^2) = N_r\} N_r.$$

Let a be a letter that occurs in $(u_{n^2})_{n\in\mathbb{N}}$. By (5.2) there exists a vector z depending on a such that

$$z^T T(v) e_1 = \begin{cases} 1, & \text{if } u_v = a, \\ 0, & \text{otherwise,} \end{cases}$$

for every $v \in \mathbb{N}$. Thus, we obtain

$$\begin{aligned} &\lim_{N\to\infty} \frac{1}{N} \#\{0 \leqslant n < N : u_{n^2} = a\} \\ &\qquad = \sum_{r=0}^{s-1} \lim_{N\to\infty} \frac{1}{N} \#\{0 \leqslant n < N : T(n^2) = N_r\} z^T N_r e_1. \end{aligned}$$

Theorem 4.1 implies that all limits exist, which in turn proves Theorem 5.2. □

5.3 Automatic sequences and the sequence $\lfloor n^c \rfloor$, $n \geqslant 0$

In this section, we show that the frequency of letters in the q-automatic sequence $(u_n)_{n\in\mathbb{N}}$ are linked to the frequency of letters in the subsequence $(u_{\lfloor n^c \rfloor})_{n\in\mathbb{N}}$ for $1 < c < 7/5$. Contrary to the case of subsequences of squares, the frequencies in $(u_n)_{n\in\mathbb{N}}$ and the one in its subsequence $(u_{\lfloor n^c \rfloor})_{n\in\mathbb{N}}$ cannot be different.

Theorem 5.8. *Let $c \in (1, 7/5)$, $q \geqslant 2$ and $(u_n)_{n\geqslant 0}$ be a q-automatic sequence. Furthermore, let a be a letter occurring in $(u_n)_{n\geqslant 0}$. Then we have that the logarithmic frequency of a in $(u_{\lfloor n^c \rfloor})_{n\geqslant 0}$ exists and it is the same as the logarithmic frequency of a in $(u_n)_{n\geqslant 0}$. Moreover, the frequency of a in $(u_n)_{n\geqslant 0}$ exists if and only if the frequency of a in $(u_{\lfloor n^c \rfloor})_{n\geqslant 0}$ exists.*

In order to show this theorem, we generalize a result of Mauduit and Rivat [MR05, Theorem 1] (see also [MR95]) to multidimensional q-multiplicative functions. They have shown that for all q-multiplicative functions f the following result holds true: If $c \in (1, 7/5)$, $\gamma = 1/c$ and $q \geqslant 2$, then for all $\delta \in (0, (7-5c)/9)$

$$\left| \sum_{1 \leqslant n \leqslant x} f(\lfloor n^c \rfloor) - \sum_{1 \leqslant m \leqslant x^c} \gamma m^{\gamma-1} f(m) \right|_s \ll x^{1-\delta},$$

where the implied constant may depend on c, δ and d. Recall that a q-multiplicative function $f : \mathbb{N} \to \mathbb{C}$ is defined by the property that for every triple $(a, b, k) \in \mathbb{N}^3$ with $b < q^k$ we have

$$f(q^k a + b) = f(q^k a) f(b).$$

We need the following definition:

Definition 5.9. Let $d \geqslant 1$. We call a function $F : \mathbb{N} \to \mathbb{C}^{d \times d}$ a generalized q-multiplicative function in $\mathbb{C}^{d \times d}$, if for every triple $(a, b, k) \in \mathbb{N}^3$ with $a > 0$ and $b < q^k$ we have

$$F(q^k a + b) = F(b) F(0)^{k - \ell(b)} F(a),$$

where $\ell(b) = \lfloor \log_q(b) \rfloor + 1$ for $b \geqslant 1$ and $\ell(0) = 1$.

The expression $\ell(b)$ is equal to the number of digits of b in the base-q representation system. If $d = 1$ and $F(0) = 1$, then F is a completely q-multiplicative function (in the classical sense). The function $S(n)$ defined in (5.1) is an important example of a generalized q-multiplicative function. Note, that we do not ask $F(0)$ to be the identity element in $\mathbb{C}^{d \times d}$ (and indeed, $S(0)$ also needs not to be the identity element).

Theorem 5.10. *Let $c \in (1, 7/5)$, $q \geqslant 2$, $d \geqslant 1$ and assume that F is a generalized q-multiplicative function in $\mathbb{C}^{d \times d}$ and there exists a submultiplicative norm $\| \cdot \|_s$ such that $\|F(k)\|_s \leqslant 1$ for all $0 \leqslant k < q$. Set $\gamma = 1/c$. Then we have for all $\delta \in (0, (7-5c)/9)$ that*

$$\left\| \sum_{1 \leqslant n \leqslant x} F(\lfloor n^c \rfloor) - \sum_{1 \leqslant m \leqslant x^c} \gamma m^{\gamma-1} F(m) \right\|_s \ll x^{1-\delta},$$

where the implied constant may depend on c, δ and d.

5.3.1 Proof of Theorem 5.10

The proof of this theorem goes along the line of Mauduit's and Rivat's proof of [MR05, Theorem 1]. In particular, many steps of their method work also in our setting. Nevertheless, we give a detailed proof for the readers convenience and refer only in few places to the work of Mauduit and Rivat.

Let $\| \cdot \|_s$ be the norm mentioned in Theorem 5.10. We denote by $\| \cdot \|_{\max}$ the maximum norm (*i.e.*, if $A = (a_{ij}) \in \mathbb{C}^{d \times d}$, then $\|A\|_{\max} = \max_{i,j} |a_{ij}|$) that is not submultiplicative.

The first steps of the proof are analog to [MR05]. The only difference is the fact that we use the triangle inequality for arbitrary norms in $\mathbb{C}^{d\times d}$ instead of the triangle inequality for the absolute value in $\mathbb{C}$. Recall that $c > 1$. A short calculation shows that m has the form $m = \lfloor n^c \rfloor$ if and only if

$$\lfloor -m^\gamma \rfloor - \lfloor -(m+1)^\gamma \rfloor = 1,$$

where $\gamma = 1/c$ (otherwise, this difference is zero). If we set $\Psi(u) = u - \lfloor u \rfloor - 1/2$, then we obtain

$$\begin{aligned}\sum_{1\leqslant n\leqslant x} F(\lfloor n^c \rfloor) &= \sum_{1\leqslant m\leqslant x^c} F(m)\left(\lfloor -m^\gamma \rfloor - \lfloor -(m+1)^\gamma \rfloor\right)\\ &= \sum_{1\leqslant m\leqslant x^c} F(m)\left((m+1)^\gamma - m^\gamma\right) \qquad (5.3)\\ &\qquad + \sum_{1\leqslant m\leqslant x^c} F(m)\left(\Psi(-(m+1)^\gamma) - \Psi(-m^\gamma)\right).\end{aligned}$$

Next, we recall a result which can be found in [MR05, Lemma 2]. If $\theta \in [0,1]$, then

$$\sum_{m\geqslant 1} \left|(m+1)^\theta - m^\theta - \theta m^{\theta-1}\right| \leqslant \frac{1}{4}. \qquad (5.4)$$

Note, that we have

$$\|F(m)\|_s \leqslant 1 \qquad (5.5)$$

for all $m \in \mathbb{N}$. This follows from the fact that F is a generalized q-multiplicative function in $\mathbb{C}^{d\times d}$, $\|\cdot\|_s$ is submultiplicative and $\|F(k)\|_s \leqslant 1$ for $0 \leqslant k < q$. Using (5.4) and (5.5), we get

$$\left\| \sum_{1\leqslant m\leqslant x^c} F(m)\left((m+1)^\gamma - m^\gamma\right) - \sum_{1\leqslant m\leqslant x^c} \gamma m^{\gamma-1} F(m) \right\|_s \leqslant \frac{1}{4}.$$

Together with (5.3), this implies that in order to prove Theorem 5.10 it suffices to show that for all $\delta \in (0, (7-5c)/9)$ and for all $M \gg 1$ we have

$$S_M := \left\| \sum_{M<m\leqslant 2M} F(m)\left(\Psi(-(m+1)^\gamma) - \Psi(-m^\gamma)\right) \right\|_s \ll M^{\gamma(1-\delta)}, \qquad (5.6)$$

where the implied constant depends on c, δ and d. Indeed, setting $M_k = x^c/2^k$, we have

$$\begin{aligned}&\left\| \sum_{1\leqslant n\leqslant x} F(\lfloor n^c \rfloor) - \sum_{1\leqslant m\leqslant x^c} \gamma m^{\gamma-1} F(m) \right\|_s\\ &\qquad \leqslant \sum_{k\geqslant 0} \left\| \sum_{M_{k+1}<m\leqslant M_k} F(m)\left(\Psi(-(m+1)^\gamma) - \Psi(-m^\gamma)\right) \right\|_s + \frac{1}{4}\\ &\qquad \ll \sum_{k\geqslant 0} M_{k+1}^{\gamma(1-\delta)} \ll x^{1-\delta}.\end{aligned}$$

We start with approximating the function Ψ by trigonometric polynomials. Let $H \geqslant 1$ be an integer. Then it follows from Proposition A.20 (Vaaler's method) that the following holds true: There exist coefficients $a_H(h)$ with $0 \leqslant a_H(h) \leqslant 1$ such that the trigonometric polynomials

$$\Psi_H(t) = -\frac{1}{2i\pi} \sum_{1 \leqslant |h| \leqslant H} \frac{a_H(h)}{h} \, \mathrm{e}(ht)$$

and

$$\kappa_H(t) = \sum_{|h| \leqslant H} \left(1 - \frac{|h|}{H+1}\right) \mathrm{e}(ht) \tag{5.7}$$

verify

$$|\Psi(t) - \Psi_H(t)| \leqslant \frac{1}{2H+2} \kappa_H(t).$$

Recall, that we used the function $\kappa_H(t)$ already in Chapter 2. We have

$$\frac{1}{2H+2} \sum_{M \leqslant m \leqslant 2M} \kappa_H(m^\theta) \ll_\theta H^{-1}M + H^{1/2}M^{\theta/2} + H^{-1/2}M^{1-\theta/2}, \tag{5.8}$$

for every $0 < \theta < 1$ and for every $M \geqslant 1$ (this is [MR05, Lemma 5] and follows easily from Theorem A.13). We set $H_0 := \lfloor M^{1-\gamma(1-\delta)} \rfloor$, where we choose $\delta > 0$ later on and obtain

$$\begin{aligned} S_M \leqslant & \left\| \sum_{M < m \leqslant 2M} F(m) \left(\Psi_{H_0}(-(m+1)^\gamma) - \Psi_{H_0}(-m^\gamma) \right) \right\|_s \\ & + \frac{1}{2H_0+2} \sum_{M < m \leqslant 2M} \kappa_{H_0}\left(-(m+1)^\gamma\right) + \frac{1}{2H_0+2} \sum_{M < m \leqslant 2M} \kappa_{H_0}\left(-m^\gamma\right). \end{aligned}$$

The last two sums can be handled by (5.8). This yields

$$\begin{aligned} S_M \ll_c & \left\| \sum_{M < m \leqslant 2M} F(m) \left(\Psi_{H_0}(-(m+1)^\gamma) - \Psi_{H_0}(-m^\gamma) \right) \right\|_s \\ & + H_0^{-1}M + H_0^{1/2}M^{\gamma/2} + H_0^{-1/2}M^{1-\gamma/2}. \end{aligned}$$

For our special choice of H_0 we have that

$$H_0^{1/2}M^{\gamma/2} = M^{1/2+\gamma\delta/2} \geqslant M^{1/2-\gamma\delta/2} = H_0^{-1/2}M^{1-\gamma/2}.$$

Thus we get

$$\begin{aligned} S_M \ll_c & \left\| \sum_{M < m \leqslant 2M} F(m) \left(\Psi_{H_0}(-(m+1)^\gamma) - \Psi_{H_0}(-m^\gamma) \right) \right\|_s \\ & + M^{\gamma(1-\delta)} + M^{1/2+\gamma\delta/2}. \end{aligned} \tag{5.9}$$

Next, we treat the sum that arises in (5.9). Replacing Ψ_{H_0} by its expression, this sum is bounded above by

$$\sum_{1\leqslant|h|\leqslant H_0} \frac{1}{|h|} \left\| \sum_{M<m\leqslant 2M} F(m)\left(\mathrm{e}\left(h(m+1)^\gamma\right) - \mathrm{e}\left(hm^\gamma\right)\right) \right\|_s$$

$$= \sum_{\ell\geqslant 0} \sum_{H_{\ell+1}<|h|\leqslant H_\ell} \frac{1}{|h|} \left\| \sum_{M<m\leqslant 2M} F(m)\left(\mathrm{e}\left(h(m+1)^\gamma\right) - \mathrm{e}\left(hm^\gamma\right)\right) \right\|_s, \qquad (5.10)$$

where $H_\ell = H_0/2^\ell$. Putting $\varphi_h(t) = \mathrm{e}(h(t+1)^\gamma - ht^\gamma) - 1$, we get by partial summation[1]

$$\sum_{M<m\leqslant 2M} F(m)\left(\mathrm{e}\left(h(m+1)^\gamma\right) - \mathrm{e}\left(hm^\gamma\right)\right)$$

$$= \varphi_h\left(2M\right) \sum_{M<m\leqslant 2M} F(m)\,\mathrm{e}\left(hm^\gamma\right) - \int_M^{2M} \varphi_h'(t) \sum_{M<m\leqslant t} F(m)\,\mathrm{e}\left(hm^\gamma\right)\mathrm{d}t.$$

If $|h| \leqslant M^{1-\gamma}$ we have $\varphi_h(t) \ll |h|M^{\gamma-1}$ and $\varphi_h'(t) \ll |h|M^{\gamma-2}$ on the interval $[M, 2M]$. Thus we obtain for $H_\ell \leqslant M^{1-\gamma}$ (note, that $\|\int A(t)\,\mathrm{d}t\|_{\max} \leqslant \int \|A(t)\|_{\max}\mathrm{d}t$ which implies $\|\int A(t)\,\mathrm{d}t\|_s \ll_d \int \|A(t)\|_s\mathrm{d}t$)

$$\sum_{H_{\ell+1}<|h|\leqslant H_\ell} \frac{1}{|h|} \left\| \sum_{M<m\leqslant 2M} F(m)\left(\mathrm{e}\left(h(m+1)^\gamma\right) - \mathrm{e}\left(hm^\gamma\right)\right) \right\|_s$$

$$\ll_d \max_{M'\in[M,2M]} M^{\gamma-1} \sum_{H_{\ell+1}<|h|\leqslant H_\ell} \left\| \sum_{M<m\leqslant M'} F(m)\,\mathrm{e}\left(hm^\gamma\right) \right\|_s.$$

Moreover, we trivially get for $\ell \geqslant 0$,

$$\sum_{H_{\ell+1}<|h|\leqslant H_\ell} \frac{1}{|h|} \left\| \sum_{M<m\leqslant 2M} F(m)\left(\mathrm{e}\left(h(m+1)^\gamma\right) - \mathrm{e}\left(hm^\gamma\right)\right) \right\|_s$$

$$\ll \max_{u\in\{0,1\}} \frac{1}{H_{\ell+1}} \sum_{H_{\ell+1}<|h|\leqslant H_\ell} \left\| \sum_{M<m\leqslant 2M} F(m)\,\mathrm{e}\left(h(m+u)^\gamma\right) \right\|_s.$$

Since the sum over ℓ in (5.10) has $\ll \log(H_0)$ summands, we obtain

$$\left\| \sum_{M<m\leqslant 2M} F(m)\left(\Psi_{H_0}(-(m+1)^\gamma) - \Psi_{H_0}(-m^\gamma)\right) \right\|_s$$

$$\ll_{c,d} (\log H_0) \max_{0<H\leqslant H_0} \max_{u\in\{0,1\}} \max_{\tilde{M}\in[M,2M]} \min\left(M^{1-\gamma}, H^{-1}\right) S(H, M, M', u), \qquad (5.11)$$

[1] If $A(t) \in \mathbb{C}^{d\times d}$, we denote by $\int A(t)\mathrm{d}t$ the matrix (B_{ij}) with $B_{ij} = \int A_{ij}(t)\mathrm{d}t$.

where $S(H, M, \tilde{M}, u)$ is defined by

$$S(H, M, M', u) = \sum_{H<h\leqslant 2H} \left\| \sum_{M<m\leqslant M'} F(m)\,\mathrm{e}\left(h(m+u)^{\gamma}\right) \right\|_s . \tag{5.12}$$

Proposition 5.11. *Let* $\gamma \in (1/2, 1)$*,* $q \geqslant 2$*,* $d \geqslant 1$ *and* F *be given as in Theorem 5.10. Then we have for all* $1/2 \leqslant H \leqslant M \leqslant M' \leqslant 2M$ *and* $u \in [0,1]$ *that*

$$\sum_{H<h\leqslant 2H} \left\| \sum_{M<m\leqslant M'} F(m)\,\mathrm{e}\left(h(m+u)^{\gamma}\right) \right\|_s \\ \ll H^{9/8} M^{(2+\gamma)/4} (1 + H^{-1/2} M^{(1-\gamma)/2}) \sqrt{\log(3M)},$$

where the implied constant depends on γ *and* d*.*

As in [MR05, page 195] one can now show that this result implies (5.6). This in turn (as already noted) proves Theorem 5.10. Thus, we omit the details and continue with proving Proposition 5.11. It turns out that it is possible to proceed similar to Mauduit and Rivat. Since the next few steps are of particular importance, we treat them in detail. The final steps are again as in [MR05, Section 4], see the comments at the end of the proof of Proposition 5.11.

Proof of Proposition 5.11. Let k be a positive integer such that[2] $B \asymp H^{-1/4} M^{1-\gamma/2}$ with $B = q^k$. We can assume that $k \geqslant 1$ (otherwise, the statement holds trivially). Then there exist integers A, R, A' and R' such that

$$M = AB + R \text{ with } 0 \leqslant R < B, \quad \text{and} \quad M' = A'B + R' \text{ with } 0 \leqslant R' < B.$$

We have $A \leqslant A' \leqslant 2A + 1$ and $AB \asymp M$. This allows us to write

$$S(H, M, M', u) = \sum_{H<h\leqslant 2H} \left\| \sum_{A\leqslant a<A'} \sum_{0\leqslant b<B} F(Ba+b)\,\mathrm{e}\left(h(Ba+b+u)^{\gamma}\right) \right\|_s + O(HB).$$

Next, Taylor's theorem implies that

$$e(h(Ba+b+u)\gamma) = \mathrm{e}(hB^{\gamma}a^{\gamma})\,\mathrm{e}(\mathbf{x}(a,h)\cdot\mathbf{y}(b)) + O_{\gamma}(HB^4M^{\gamma-4}),$$

where

$$\mathbf{x}(a,h) = (ha^{\gamma-1}, ha^{\gamma-2}, ha^{\gamma-3}), \\ \mathbf{y}(b) = (\gamma_1 B^{\gamma-1}(b+u), \gamma_2 B^{\gamma-2}(b+u)^2, \gamma_3 B^{\gamma-3}(b+u)^3),$$

with $\gamma_1 = \gamma$, $\gamma_2 = \gamma(\gamma-1)/2$ and $\gamma_3 = \gamma(\gamma-1)\gamma-2)/6$. Thus, we have

$$S(H, M, M', u) = \sum_{H<h\leqslant 2H} \left\| \sum_{A\leqslant a<A'} \sum_{0\leqslant b<B} F(Ba+b)\,\mathrm{e}(hB^{\gamma}a^{\gamma})\,\mathrm{e}(\mathbf{x}(a,h)\cdot\mathbf{y}(b)) \right\|_s \\ + O(HB + H^2B^4M^{\gamma-3}).$$

[2] The symbol $f \asymp g$ means that $f \ll g \ll f$.

Let us define the function $\tilde{F} : \mathbb{N} \to \mathbb{C}^{d\times d}$ by

$$\tilde{F}(n) = \begin{cases} F(n)F(0)^{k-\ell(n)}, & \text{if } n < q^k,\\ 0^{d\times d}, & \text{otherwise,}\end{cases}$$

where $0^{d\times d}$ is the zero matrix in $\mathbb{C}^{d\times d}$. Recall, that $\ell(n)$ denotes the number of digits of n. The generalized q-multiplicity of F implies that we have for all $A \leqslant a < A'$ and $0 \leqslant b < B$ that

$$F(Ba+b) = \tilde{F}(b)F(a).$$

We obtain

$$\begin{aligned} S(H,M,M',u) &= \sum_{H<h\leqslant 2H} \left\| \sum_{A\leqslant a<A'} \sum_{0\leqslant b<B} \tilde{F}(b)F(a)\,\mathrm{e}(hB^\gamma a^\gamma)\,\mathrm{e}(\mathbf{x}(a,h)\cdot\mathbf{y}(b)) \right\|_s \\ &\qquad + O(HB + H^2B^4M^{\gamma-3}) \\ &\ll \sum_{H<h\leqslant 2H} \sum_{A\leqslant a<A'} \left\| \sum_{0\leqslant b<B} \tilde{F}(b)\,\mathrm{e}(\mathbf{x}(a,h)\cdot\mathbf{y}(b)) \right\|_s \\ &\qquad + HB + H^2B^4M^{\gamma-3}, \end{aligned}$$

where we used the submultiplicity of $\|\cdot\|_s$. Note, that $\|A\|_s \ll_d \sum_{1\leqslant i,j\leqslant d} |a_{ij}|$ for any matrix $A = (a_{ij})$. Hence we get

$$\begin{aligned} S(H,M,M',u) \ll_d &\sum_{1\leqslant i,j\leqslant d} \sum_{H<h\leqslant 2H} \sum_{A\leqslant a<A'} \left| \sum_{0\leqslant b<B} \tilde{F}_{ij}(b)\,\mathrm{e}(\mathbf{x}(a,h)\cdot\mathbf{y}(b)) \right| \\ &\qquad + HB + H^2B^4M^{\gamma-3}, \end{aligned}$$

where $\tilde{F} = (\tilde{F}_{ij})_{1\leqslant i,j\leqslant d}$. Set $\mathcal{X} = \{\mathbf{x}(a,h) : A \leqslant a < A', H < h \leqslant 2H\}$ and $\mathcal{Y} = \{\mathbf{y}(b) : 0 \leqslant b < B\}$. Note, that $\mathbf{x}(a,h) \neq \mathbf{x}(a',h')$ if $(a,h) \neq (a',h')$ and $\mathbf{y}(b) \neq \mathbf{y}(b')$ if $b \neq b'$. We obtain that there exist complex numbers $\alpha_{ij}(\mathbf{x}(a,h))$ and $\beta_{ij}(\mathbf{y}(b))$ with $|\alpha_{ij}(\mathbf{x}(a,h))| = 1$ and $|\beta_{ij}(\mathbf{y}(b))| \ll_d 1$ such that

$$\begin{aligned} &\sum_{H<h\leqslant 2H} \sum_{A\leqslant a<A'} \left| \sum_{0\leqslant b<B} \tilde{F}_{ij}(b)\,\mathrm{e}(\mathbf{x}(a,h)\cdot\mathbf{y}(b)) \right| \\ &\qquad = \sum_{\substack{\mathbf{x}(a,h)\in\mathcal{X}\\ \mathbf{y}(b)\in\mathcal{Y}}} \alpha_{ij}(\mathbf{x}(a,h))\beta_{ij}(\mathbf{y}(b))\,\mathrm{e}(\mathbf{x}(a,h)\cdot\mathbf{y}(b)). \end{aligned}$$

(As in (5.5) we see that $\|\tilde{F}(n)\|_s \leqslant 1$ for all $n \in \mathbb{N}$. This implies that $|F_{ij}(b)| \leqslant \|F(b)\|_{\max} \ll_d \|\tilde{F}(n)\|_s \ll_d 1$.) We set

$$\begin{aligned} \Delta_k^{-1} &= \gamma_k HB^kM^{\gamma-k},\\ X_k &= \gamma_k^{-1}\Delta_k^{-1}B^{-\gamma}\\ Y_k &= \Delta_k^{-1}H^{-1}A^{k-\gamma} \end{aligned}$$

for $k = 1, \ldots, 3$. Then we have that the k-th component of $\mathbf{x}(a,h) \in \mathcal{X}$ has absolute value less than or equal to X_k. A similar result holds for the points in $\mathcal{Y}$ (with X_k replaced by Y_k). Hence, we can apply Theorem A.17 (the double large sieve of Bombieri and Iwaniec) with $\kappa = 3$ and obtain

$$\left(\sum_{H<h\leqslant 2H} \sum_{A\leqslant a<A'} \left| \sum_{0\leqslant b<B} \tilde{F}_{ij}(b)\, \mathrm{e}(\mathbf{x}(a,h)\cdot \mathbf{y}(b)) \right| \right)^2 \ll_d \prod_{k=1}^{3} (1+\Delta_k^{-1}) \mathcal{B}_1 \mathcal{B}_2, \quad (5.13)$$

where $\mathcal{B}_1$ represents the number of quadruples (h_1, h_2, a_1, a_2) with $H \leqslant h_1, h_2 \leqslant 2H$ and $A \leqslant a_1, a_2 \leqslant A'$ such that

$$\left| h_1 a_1^{\gamma-k} - h_2 a_2^{\gamma-k} \right| \leqslant (2Y_k)^{-1}, \qquad k = 1, \ldots, 3,$$

and $\mathcal{B}_2$ represents the number of pairs (b_1, b_2) with $0 \leqslant b_1, b_2 < B$ such that

$$\left| \gamma_k B^{\gamma-k}(b_1+u)^k - \gamma_k B^{\gamma-k}(b_2+u)^k \right| \leqslant (2X_k)^{-1}, \qquad k = 1, \ldots, 3.$$

Note, that the right-hand side of (5.13) is independent of i and j, since the sets $\mathcal{X}$ and $\mathcal{Y}$ as well as the numbers X_k and Y_k for $k = 1, \ldots, 3$ are independent of i and j. Mauduit and Rivat have shown (see [MR05, Sections 3 and 4]) that

$$\prod_{k=1}^{3} (1+\Delta_k^{-1}) \mathcal{B}_1 \mathcal{B}_2 \ll_\gamma H^{9/4} M^{1+\gamma/2} (1 + H^{-1} M^{1-\gamma}) \log(3M).$$

Thus, we obtain

$$\begin{aligned} S(H,M,M',u) \ll_{\gamma,d} & \; H^{9/8} M^{(2+\gamma)/4} (1 + H^{-1/2} M^{(1-\gamma)/2}) \sqrt{\log(3M)} \\ & + HB + H^2 B^4 M^{\gamma-3}. \end{aligned} \quad (5.14)$$

Exactly the same way as at the end of [MR05, Sections 4], we obtain the the last two terms on the right-hand side of (5.14) are smaller than the first term. Finally, this completes the proof of Proposition 5.11. □

5.3.2 Proof of Theorem 5.8

Before we start with proving Theorem 5.8, we show that Theorem 5.10 implies the following result.

Lemma 5.12. *Let $c \in (1, 7/5)$, $q \geqslant 2$, $d \geqslant 1$ and assume that F is a generalized q-multiplicative function in $\mathbb{C}^{d\times d}$ and there exists a submultiplicative norm $\|\cdot\|_s$ such that $\|F(k)\|_s \leqslant 1$ for all $0 \leqslant k < q$. Set $\gamma = 1/c$. Then we have*

$$\left\| \sum_{1\leqslant n\leqslant x} \frac{F(\lfloor n^c \rfloor)}{n} - \sum_{1\leqslant m\leqslant x^c} \gamma \frac{F(m)}{m} \right\|_s \ll_{c,d} 1, \quad (5.15)$$

and for all $\delta \in (0, (7-5c)/9)$,

$$\left\| \sum_{1 \leqslant n \leqslant x^\gamma} c n^{c-1} F(\lfloor n^c \rfloor) - \sum_{1 \leqslant m \leqslant x} F(m) \right\|_s \ll_{c,\delta,d} x^{1-\delta\gamma}. \tag{5.16}$$

Proof. We start with proving inequality (5.15). By partial summation we obtain

$$\begin{aligned} &\sum_{1 \leqslant n \leqslant x} \frac{F(\lfloor n^c \rfloor)}{n} - \sum_{1 \leqslant m \leqslant x^c} \gamma \frac{F(m)}{m} \\ &\qquad = \frac{1}{x} \sum_{1 \leqslant n \leqslant x} F(\lfloor n^c \rfloor) - \frac{1}{x} \sum_{1 \leqslant m \leqslant x^c} \gamma m^{\gamma-1} F(m) + I(x), \end{aligned}$$

where

$$I(x) = \int_1^x \left(\sum_{1 \leqslant n \leqslant t} F(\lfloor n^c \rfloor) \right) \frac{1}{t^2} \mathrm{d}t - \gamma \int_1^{x^c} \left(\sum_{1 \leqslant m \leqslant t} \gamma m^{\gamma-1} F(m) \right) \frac{1}{t^{\gamma+1}} \mathrm{d}t.$$

Changing the variable in the last integral yields

$$I(x) = \int_1^x \left(\sum_{1 \leqslant n \leqslant t} F(\lfloor n^c \rfloor) - \sum_{1 \leqslant m \leqslant t^c} \gamma m^{\gamma-1} F(m) \right) \frac{1}{t^2} \mathrm{d}t.$$

Thus we obtain

$$\begin{aligned} &\left\| \sum_{1 \leqslant n \leqslant x} \frac{F(\lfloor n^c \rfloor)}{n} - \sum_{1 \leqslant m \leqslant x^c} \gamma \frac{F(m)}{m} \right\|_s \\ &\qquad \ll_d \frac{1}{x} \left\| \sum_{1 \leqslant n \leqslant x} F(\lfloor n^c \rfloor) - \sum_{1 \leqslant m \leqslant x^c} \gamma m^{\gamma-1} F(m) \right\|_s \\ &\qquad\qquad + \int_1^x \left\| \sum_{1 \leqslant n \leqslant t} F(\lfloor n^c \rfloor) - \sum_{1 \leqslant m \leqslant t^c} \gamma m^{\gamma-1} F(m) \right\|_s \frac{1}{t^2} \mathrm{d}t. \end{aligned}$$

We can use Theorem 5.10 with some fixed $\delta < (7-5c)/9$ and obtain

$$\left\| \sum_{1 \leqslant n \leqslant x} \frac{F(\lfloor n^c \rfloor)}{n} - \sum_{1 \leqslant m \leqslant x^c} \gamma \frac{F(m)}{m} \right\|_s \ll_{c,d} \frac{1}{x^\delta} + \int_1^x \frac{1}{t^{1+\delta}} \mathrm{d}t \ll_{c,d} 1.$$

The proof of (5.16) uses the same ideas. Partial summation yields

$$\begin{aligned} &\sum_{1 \leqslant n \leqslant x^\gamma} n^{c-1} F(\lfloor n^c \rfloor) - \sum_{1 \leqslant m \leqslant x} \gamma F(m) \\ &\qquad = x^{1-\gamma} \sum_{1 \leqslant n \leqslant x^\gamma} F(\lfloor n^c \rfloor) - x^{1-\gamma} \sum_{1 \leqslant m \leqslant x} \gamma m^{\gamma-1} F(m) - J(x), \end{aligned}$$

where

$$J(x) = (c-1)\int_1^{x^\gamma}\left(\sum_{1\leqslant n\leqslant t} F(\lfloor n^c\rfloor)\right)\frac{1}{t^{2-c}}\mathrm{d}t$$
$$+ (1-\gamma)\int_1^{x}\left(\sum_{1\leqslant m\leqslant t} \gamma m^{\gamma-1}F(m)\right)\frac{1}{t^{\gamma}}\mathrm{d}t.$$

Integration by substitution implies

$$\left\|\sum_{1\leqslant n\leqslant x^\gamma} n^{c-1}F(\lfloor n^c\rfloor) - \sum_{1\leqslant m\leqslant x}\gamma F(m)\right\|_s$$
$$\ll_d x^{1-\gamma}\left\|\sum_{1\leqslant n\leqslant x^\gamma} F(\lfloor n^c\rfloor) - \sum_{1\leqslant m\leqslant x}\gamma m^{\gamma-1}F(m)\right\|_s$$
$$+ (1-\gamma)\int_1^{x}\left\|\sum_{1\leqslant n\leqslant t^\gamma} F(\lfloor n^c\rfloor) - \sum_{1\leqslant m\leqslant t}\gamma m^{\gamma-1}F(m)\right\|_s\frac{1}{t^{\gamma}}\mathrm{d}t,$$

and we finally obtain (again using Theorem 5.10)

$$\left\|\sum_{1\leqslant n\leqslant x^\gamma} n^{c-1}F(\lfloor n^c\rfloor) - \sum_{1\leqslant m\leqslant x}\gamma F(m)\right\|_s \ll_{c,\delta,d} x^{1-\gamma\delta} + \int_1^x \frac{1}{t^{\gamma\delta}}\mathrm{d}t \ll_{c,\delta,d} x^{1-\gamma\delta},$$

for any $\delta \in (0, (7-5c)/9)$. □

Proof of Theorem 5.8. Recall that we have $1 < c < 7/5$ and $q \geqslant 2$. Let a be a letter occurring in the q-automatic sequence $(u_n)_{n\geqslant 0}$ (realized by a finite automaton with d states). As we have seen in Section 5.1, there exist q transition matrices $M_0, \ldots, M_{q-1} \in \mathbb{C}^{d\times d}$ corresponding to the automatic sequence and a vector $z = z_a \in \mathbb{C}^d$ such that

$$z^T S(n) e_1 = \begin{cases} 1, & \text{if } u_n = a,\\ 0, & \text{otherwise,}\end{cases}$$

where $S(n)$ is given by (5.1). Recall, that $S(n)$ is a generalized q-multiplicative function in $\mathbb{C}^{d\times d}$. Let $\|\cdot\|_1$ denote the submultiplicative norm induced by the 1-norm in $\mathbb{C}^d$. Alternatively, if $A = (a_{ij})_{1\leqslant i,j\leqslant d}$, then $\|A\|_1$ is also given by $\|A\|_1 = \max_j \sum_i |a_{ij}|$ (maximum absolute column sum norm). Since for each n there is exactly one entry equal to 1 in each column of $S(n)$, we have $\|S(n)\|_1 = 1$.

We start with showing that the logarithmic frequency of a in $(u_{\lfloor n^c\rfloor})_{n\geqslant 0}$ exists and that it is the same as the logarithmic frequency $j_{\log}^{(a)}$ of a in $(u_n)_{n\geqslant 0}$ (which exists since this sequence is q-automatic). We have

$$j_{\log}^{(a)} = \lim_{x\to\infty} \frac{1}{\log x} \sum_{\substack{0\leqslant n<x\\ u_n=a}} \frac{1}{n} = \lim_{x\to\infty} \frac{1}{\log x} \sum_{1\leqslant n\leqslant x} \frac{z^T S(n) e_1}{n}.$$

In what follows, we show that

$$\frac{1}{\log x} \sum_{1\leqslant n\leqslant x} \frac{z^T S(\lfloor n^c \rfloor) e_1}{n} - j_{\log}^{(a)} = o(1),$$

which implies the desired result. We can write

$$\frac{1}{\log x} \sum_{1\leqslant n\leqslant x} \frac{z^T S(\lfloor n^c \rfloor) e_1}{n} - j_{\log}^{(a)} = E_{\log}^{(1)} + E_{\log}^{(2)},$$

with

$$\begin{aligned} E_{\log}^{(1)} &= \frac{1}{\log x} z^T \left(\sum_{1\leqslant n\leqslant x} \frac{S(\lfloor n^c \rfloor)}{n} - \sum_{1\leqslant m\leqslant x^c} \gamma \frac{S(m)}{m} \right) e_1, \\ E_{\log}^{(2)} &= \frac{1}{\log x^c} \sum_{1\leqslant m\leqslant x^c} \frac{z^T S(m) e_1}{m} - j_{\log}^{(a)}. \end{aligned}$$

The Cauchy-Schwarz inequality yields

$$\begin{aligned} \left| E_{\log}^{(1)} \right| &\leqslant \frac{1}{\log x} \|z\|_2 \cdot \left\| \left(\sum_{1\leqslant n\leqslant x} \frac{S(\lfloor n^c \rfloor)}{n} - \sum_{1\leqslant m\leqslant x^c} \gamma \frac{S(m)}{m} \right) e_1 \right\|_2 \\ &\leqslant \frac{\sqrt{d}}{\log x} \left\| \sum_{1\leqslant n\leqslant x} \frac{S(\lfloor n^c \rfloor)}{n} - \sum_{1\leqslant m\leqslant x^c} \gamma \frac{S(m)}{m} \right\|_2, \end{aligned} \tag{5.17}$$

where $\|\cdot\|_2$ denotes the norm induced by the 2-norm in $\mathbb{C}^d$. Inequality (5.15) of Lemma 5.12 implies that $E_{\log}^{(1)} = o(1)$. Since $E_{\log}^{(2)} = o(1)$ holds trivially, we are done.

Nest, we assume that the frequency of a in $(u_n)_{n\geqslant 0}$ exists (and we denote it by the symbol $j^{(a)}$). Then we have to show that

$$\sum_{1\leqslant n\leqslant x} \left(z^T S(\lfloor n^c \rfloor) e_1 - j^{(a)} \right) = o(x). \tag{5.18}$$

Again, we split up the occurring sum in different parts. We write

$$\sum_{1\leqslant n\leqslant x} \left(z^T S(\lfloor n^c \rfloor) e_1 - j^{(a)} \right) = E^{(1)} + E^{(2)} + E^{(3)},$$

with

$$E^{(1)} = z^T \left(\sum_{1\leqslant n\leqslant x} S(\lfloor n^c \rfloor) - \sum_{1\leqslant m\leqslant x^c} \gamma m^{\gamma-1} S(m) \right) e_1,$$

$$E^{(2)} = \sum_{1\leqslant m\leqslant x^c} \gamma m^{\gamma-1} \left(z^T S(m) e_1 - j^{(a)} \right),$$

$$E^{(3)} = j^{(a)} \left(\sum_{1\leqslant m\leqslant x^c} \gamma m^{\gamma-1} - \sum_{1\leqslant n\leqslant x} 1 \right).$$

Similar to (5.17) we get

$$\left| E^{(1)} \right| \leqslant \sqrt{d} \left\| \sum_{1\leqslant n\leqslant x} S(\lfloor n^c \rfloor) - \sum_{1\leqslant m\leqslant x^c} \gamma m^{\gamma-1} S(m) \right\|_2 .$$

Theorem 5.10 implies that $E^{(1)} = o(x)$. In order to treat $E^{(2)}$ note that there exists a continuous, positive and monotonic function $g(t)$ with $g(t) \to 0$ for $t \to \infty$ such that

$$\left| \sum_{1\leqslant m\leqslant t} \left(z^T S(m) e_1 - j^{(a)} \right) \right| \leqslant t g(t).$$

(This easily follows from the fact that the frequency of a exists.) By partial summation we obtain

$$E^{(2)} = \gamma x^{1-c} \sum_{1\leqslant m\leqslant x^c} \left(z^T S(m) e_1 - j^{(a)} \right) \\ - \int_1^{x^c} \left(\sum_{1\leqslant m\leqslant t} \left(z^T S(m) e_1 - j^{(a)} \right) \right) \gamma(\gamma-1) t^{\gamma-2} \mathrm{d}t.$$

Hence we have

$$\left| E^{(2)} \right| \leqslant \gamma x^{1-c} x^c g(x^c) + \int_1^{x^{c/2}} g(t)\gamma(1-\gamma) t^{\gamma-1} \mathrm{d}t + \int_{x^{c/2}}^{x^c} g(t)\gamma(1-\gamma) t^{\gamma-1} \mathrm{d}t \\ \ll x g(x^c) + x^{1/2} + g(x^{c/2}) x. \tag{5.19}$$

This implies $E^{(2)} = o(x)$. That $E^{(3)} = o(x)$ is a simply consequence of Euler-Maclaurin's summation formula (see for example [Ten08, Theorem 0.7]). We finally obtain that (5.18) holds true.

In oder to finish the proof of Theorem 5.8 it remains to show that existence of the frequency of a in $(u_{\lfloor n^c \rfloor})_{n\geqslant 0}$ implies existence of the frequency of a in $(u_n)_{n\geqslant 0}$. If

$$j_c^{(a)} = \lim_{x\to\infty} \frac{1}{x} \sum_{1\leqslant n\leqslant x} z^T S(\lfloor n^c \rfloor) e_1,$$

we have to show that

$$\sum_{1 \leqslant m \leqslant x} \left(z^T S(m) e_1 - j_c^{(a)} \right) = o(x).$$

We get

$$\sum_{1 \leqslant m \leqslant x} \left(z^T S(m) e_1 - j_c^{(a)} \right) = E_c^{(1)} + E_c^{(2)} + E_c^{(3)},$$

with

$$\begin{aligned}
E_c^{(1)} &= z^T \left(\sum_{1 \leqslant m \leqslant x} S(m) - \sum_{1 \leqslant n \leqslant x^\gamma} c n^{c-1} S(\lfloor n^c \rfloor) \right) e_1, \\
E_c^{(2)} &= \sum_{1 \leqslant n \leqslant x^\gamma} c n^{c-1} \left(z^T S(\lfloor n^c \rfloor) e_1 - j_c^{(a)} \right), \\
E_c^{(3)} &= j_c^{(a)} \left(\sum_{1 \leqslant n \leqslant x^\gamma} c n^{c-1} - \sum_{1 \leqslant n \leqslant x} 1 \right).
\end{aligned}$$

Similar to the considerations we have done before, we see that $E_c^{(1)} = o(x)$ (by Lemma 5.12, Inequality (5.16)). As in (5.19) we obtain $E_c^{(2)} = o(x)$ and Euler-Maclaurin's summation formula yields again $E_c^{(3)} = o(x)$. This finishes the proof of Theorem 5.8. □

Chapter 6

The sum of digits of $\lfloor n^c \rfloor$

If numbers aren't beautiful,
I don't know what is.

Paul Erdős (1913 - 1996)

Let s_q be the sum-of-digits function in base $q \geqslant 2$ and let $c > 0$ be a real number different from an integer. As we have seen in the previous chapter, there is a strong relation between the sequence $(s_q(\lfloor n \rfloor))_{n\in\mathbb{N}}$ and its subsequence $(s_q(\lfloor n^c \rfloor))_{n\in\mathbb{N}}$ for c slightly bigger than 1. The study of distribution properties becomes more involved if the subsequence gets sparser and sparser as the value c gets bigger. This problem is linked to Gelfond's third question and polynomials are replaced by similarly increasing functions. Additionally, it can be understood as an intermediate case between polynomials of different degree.

In this chapter we show that if $q \geqslant q_0(c)$, where $q_0(c)$ is a constant depending on c, then the sequence $(\alpha s_q(\lfloor n^c \rfloor))_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 if and only if α is irrational. Furthermore, we show that for $q \geqslant q_0(c)$ the sequence $(s_q(\lfloor n^c \rfloor))_{n\in\mathbb{N}}$ is well distributed in arithmetic progressions.

6.1 Introduction and main results

In this chapter q denotes an integer $\geqslant 2$ and c is a positive real number different from an integer. We use the usual notation $\mathrm{e}(x) = e^{2\pi i x}$ and $\|x\| = \min_{z\in\mathbb{Z}} |z - x|$. Furthermore, we denote by $\{x\}$ the fractional part of x (i.e., $\{x\} = x - \lfloor x \rfloor$) and by $s_q(n)$ the ordinary sum-of-digits function of n in base q.

Recall that Gelfond [Gel68] showed that if $q, m, r > 1$ and ℓ, a are integers with $(m, q-1) = 1$, then

$$\#\{n \leqslant x : n \equiv \ell \bmod r,\ s_q(n) \equiv a \bmod m\} = \frac{x}{mr} + O(x^\lambda), \tag{6.1}$$

where $\lambda < 1$ is a positive constant depending only on q and m. If one replaces the arithmetic progression $\{n \geqslant 0 : n \equiv \ell \bmod r\}$ by another sequence, then the corresponding question is in general much harder to answer. For instance, Gelfond's second problem on the sum of digits of primes was unsolved for a long time until Mauduit and Rivat [MR10] proved that $(s_q(p))$, where p ranges over all primes, is well distributed in arithmetic progressions. The treatment of the sequence $(s_q(P(n)))_{n\in\mathbb{N}}$, where $P(n)$ is a polynomial with $P(\mathbb{N}) \subseteq \mathbb{N}$, seems to be even more complex. Dartyge and Tenenbaum [DT06] showed that if $(m, q-1) = 1$, then

$$\#\{n \leqslant x : s_q(P(n)) \equiv a \bmod m\} \geqslant C x^{\min(1,2/d!)},$$

where d is the degree of the polynomial P and C is a positive constant depending on P, q and m. Drmota, Mauduit, and Rivat [DMR] considered the sequence $(s_q(P(n)))_{n\in\mathbb{N}}$ for sufficiently large bases q and Mauduit and Rivat solved the problem in the quadratic case for general $q \geqslant 2$ (see Chapter 1).

A related question is whether a Gelfond type result also holds true for the sequence $(s_q(\lfloor n^c \rfloor))_{n\in\mathbb{N}}$, where c is a real number different from an integer. It can be understood as an intermediate case between polynomials of different degree. Mauduit and Rivat gave a positive answer for $c \in (1, 4/3)$ in 1995 (see [MR95]) and for $c \in (1, 7/5)$ in 2005 (see [MR05]). They considered more generally q-multiplicative functions and used, among other tools, the double large sieve of Bombieri and Iwaniec to solve this problem (compare with Chapter 5). Their last result reads as follows:

Theorem MR. *Let $c \in (1, 7/5)$ and $q \geqslant 2$. If a and m are integers with $m \geqslant 1$, then*

$$\lim_{x\to\infty} \frac{1}{x} \# \{n \leqslant x : s_q(\lfloor n^c \rfloor) \equiv a \bmod m\} = \frac{1}{m}. \tag{6.2}$$

Furthermore, the sequence $(\alpha s_q(\lfloor n^c \rfloor))_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 if and only if α is irrational.

As pointed out by Mauduit (see [Mau01, Section II.4]), one can deduce from a result of Harman and Rivat [HR95] that (6.2) holds for almost all $c \in [1, 2)$. Indeed, if $\mathcal{A}$ is an infinite set of positive integers such that $\#\{n \leqslant x : n \in \mathcal{A}\} \gg x$, then it follows from [HR95, Theorem 3] that for almost all $c \in (1, 2)$ we have

$$\#\{n \leqslant x : \lfloor n^c \rfloor \in \mathcal{A}\} = \gamma \sum_{\substack{n \leqslant x^c \\ n \in \mathcal{A}}} n^{-1+\gamma} + o(x), \tag{6.3}$$

where $\gamma = 1/c$. Setting $\mathcal{A} = \{n \in \mathbb{N} : s_q(n) \equiv a \bmod m\}$, a refined version of Gelfond's work (cf. (6.1)) implies that $\#\{n \leqslant x : n \in \mathcal{A}\} \gg x$. Elementary discrete Fourier analysis and partial summation (similar to Section 6.4.1) allow to evaluate the occurring sum in (6.3) and we finally obtain that (6.2) holds true for almost all $c \in (1, 2)$ and for every triple of integers (a, p, m) with $q \geqslant 2$ and $m \geqslant 1$. This leads to the following conjecture which can be found in [Mau01, Conjecture 1]:

Conjecture 6.1 (Mauduit)**.** *For almost all $c > 1$ we have for every integer q and m greater than 1 and $0 \leqslant a < m$, that*

$$\lim_{x\to\infty} \frac{1}{x} \# \{n \leqslant x : s_q(\lfloor n^c \rfloor) \equiv a \bmod m\} = \frac{1}{m}. \tag{6.4}$$

The main objective of this chapter is to enlarge the range of possible real numbers c in Theorem MR for which we can show uniform distribution results (Corollary 6.4 and 6.6). We are able to deal with all positive real numbers c which are not integers but we restrict us to bases q which are not too small. It turns out that the case $c \in \mathbb{N}$ is of completely different nature. This makes it finally impossible to treat general numbers c with the methods presented in this work. In Section 6.3 we will provide a precise analysis of this problem. (See also [DMR], where Drmota, Mauduit, and Rivat use different methods to show that $(s_q(P(n)))_{n\in\mathbb{N}}$ is well distributed in arithmetic progressions if P is a polynomial with $P(\mathbb{N}) \subseteq \mathbb{N}$ and if q is big enough.) In our main theorem we study the exponential sum $\sum_n \mathrm{e}(\alpha s_q(\lfloor n^c \rfloor))$:

Theorem 6.2. *Let $c > 0$ be a real number different from an integer and $\alpha \in \mathbb{R}$. Then there exists a constant $q_0(c)$ such that for all $q \geqslant q_0(c)$ we have*

$$\sum_{1\leqslant n\leqslant x} \mathrm{e}\left(\alpha s_q(\lfloor n^c \rfloor)\right) \ll_{c,q} (\log x)\, x^{1-\sigma_{c,q}\|(q-1)\alpha\|^2}, \tag{6.5}$$

where $\sigma_{c,q} > 0$ is a computable positive constant. In the case $0 < c < 1$ we have $q_0(c) = 2$ and the exponent in the right-hand side of (6.5) *can be replaced by* $(1 - \sigma_{c,q}\|\alpha\|^2)$.

Remark 6.3. It follows from the reasoning of our proof that an admissible value of $q_0(c)$ is explicitly computable and that this value is bounded by Kc^{c^4}, where K is an absolute constant. We use different methods to show the result for different values of c in order to optimize $q_0(c)$ (see Sections 6.2 and 6.3 and the end of Section 6.4). If $1 < c < 7/5$, then [MR05, Theorem 1] and partial summation ensures that we can choose $q_0(c) = 2$. The case $0 < c < 1$ can be seen as trivial but for the sake of completeness we give a short proof in Section 6.4.

Corollary 6.4. *Let $c > 0$ be a real number different from an integer. There exists a constant $q_0(c) \geqslant 2$ such that for all $q \geqslant q_0(c)$ the following holds: If $a, m \in \mathbb{Z}$ with $m \geqslant 1$, then there exists a constant $\sigma_{q,m,c} > 0$, such that*

$$\#\left\{n \leqslant x : s_q\left(\lfloor n^c \rfloor\right) \equiv a \bmod m\right\} = \frac{x}{m} + O_{c,q,m}\left(x^{1-\sigma_{q,m,c}}\right).$$

Remark 6.5. Corollary 6.4 does not solve Conjecture 6.1 entirely, but it leads us to conjecture that (6.4) is valid for every $c > 1$ ($c \notin \mathbb{N}$). If $c > 1$ is an integer, then elementary arithmetic calculations may yield a different asymptotic formula which depends on a, m and q (cf. [DMR]).

Corollary 6.6. *Let $c > 0$ be a real number different from an integer. Then there exists a constant $q_0(c)$, such that for all $q \geqslant q_0(c)$ the sequence $(\alpha s_q(\lfloor n^c \rfloor))_{n\in\mathbb{N}}$ is uniform distributed modulo 1 if and only if α is irrational.*

Remark 6.7. Since the estimate (6.5) is uniform in α, Theorem 6.2 (together with results from Sections 6.2 and 6.3) allows to derive a local limit theorem, see Chapter 7.

The main idea of showing Theorem 6.2 is to divide the proof up into a Fourier theory part and an exponential sums part (where no sum-of-digits function occurs). Let $q \geqslant 2$, $\alpha \in \mathbb{R}$ and $\lambda \in \mathbb{N}$. The discrete Fourier transform $F_\lambda(.,\alpha)$ of the function $u \mapsto \mathrm{e}(\alpha s_q(u))$ is defined for all $h \in \mathbb{Z}$ by

$$F_\lambda(h,\alpha) = \frac{1}{q^\lambda} \sum_{0 \leqslant u < q^\lambda} \mathrm{e}\left(\alpha s_q(u) - huq^{-\lambda}\right). \tag{6.6}$$

See Appendix A.2 for properties of this and related functions proved by Fouvry, Mauduit, and Rivat.

In the next two sections we discuss the sum $\sum_n \mathrm{e}(\beta\lfloor n^c \rfloor)$. In particular, we present a method which works for $1 < c < 2$ in Section 6.2 and a method for general real numbers $c \notin \mathbb{N}$ in Section 6.3. In Section 6.4 we prove Theorem 6.2. Finally, Section 6.5 is devoted to the proofs of Corollary 6.4 and Corollary 6.6.

6.2 Exponential sums for $1 < c < 2$

In this section we treat the exponential sum $\sum_n \mathrm{e}(\beta \lfloor n^c \rfloor)$ for $1 < c < 2$.

Proposition 6.8. *Let $1 < c < 2$ and $x, \nu \in \mathbb{N}$ with $q^{\nu-1} < x \leqslant q^\nu$. Furthermore, let $\beta \in \mathbb{R} \setminus \mathbb{Z}$. Then we have*

$$\sum_{q^{\nu-1} < n \leqslant x} \mathrm{e}\left(\beta \lfloor n^c \rfloor\right) \ll_{c,q} \nu q^{\nu(1-(2-c)/3)} + \frac{1}{\|\beta\|} q^{\nu(1-c)}.$$

The method for the proof of this proposition is based on a work of Mauduit and Rivat [MR05]. The first steps are similar to the one presented in the previous chapter (see Section 5.3.1). We use the fact that an integer m has the form $m = \lfloor n^c \rfloor$ if and only if

$$\lfloor -m^\gamma \rfloor - \lfloor -(m+1)^\gamma \rfloor = 1,$$

where $\gamma = 1/c$. If we set $\Psi(u) = u - \lfloor u \rfloor - 1/2$, then we obtain

$$\begin{aligned}\sum_{q^{\nu-1} < n \leqslant x} \mathrm{e}\left(\beta \lfloor n^c \rfloor\right) &= \sum_{q^{(\nu-1)c} < m \leqslant x^c} \mathrm{e}\left(\beta m\right) \left(\lfloor -m^\gamma \rfloor - \lfloor -(m+1)^\gamma \rfloor\right) \\ &= \sum_{q^{(\nu-1)c} < m \leqslant x^c} \mathrm{e}\left(\beta m\right) \left((m+1)^\gamma - m^\gamma\right) \qquad (6.7)\\ &\quad + \sum_{q^{(\nu-1)c} < m \leqslant x^c} \mathrm{e}\left(\beta m\right) \left(\Psi(-(m+1)^\gamma) - \Psi(-m^\gamma)\right).\end{aligned}$$

The first sum that appears on the right-hand side of (6.7) can be estimated by partial summation (see Lemma 6.9). This leads us to consider the double sum

$$S(K,\tilde{M},u) = \sum_{K < |k| \leqslant 2K} \left| \sum_{M < m \leqslant \tilde{M}} f(m)\, \mathrm{e}\left(k(m+u)^\gamma\right) \right|, \tag{6.8}$$

where $f(m) = \mathrm{e}(\beta m)$ (compare with (5.12)). The main difference to the work of Mauduit and Rivat is the fact that they have to deal with q-multiplicative functions $f(m)$ instead of $\mathrm{e}(\beta m)$. Using van der Corput's method of estimating exponential sums finally enables us to obtain the desired result (see Lemma 6.10).

Lemma 6.9. *Let $c > 1$ and $\gamma = 1/c$. Furthermore, let $x, \nu \in \mathbb{N}$ with $q^{\nu-1} < x \leqslant q^\nu$ and $\beta \in \mathbb{R} \setminus \mathbb{Z}$. Then we have*

$$\begin{aligned}\sum_{q^{(\nu-1)c} < m \leqslant x^c} \mathrm{e}(\beta m)\left((m+1)^\gamma - m^\gamma\right) &\leqslant \frac{\gamma}{|\sin \pi\beta|}\, q^{(\nu-1)(1-c)}\left(2 - q^{1-c}\right) + \frac{1}{4} \\ &\ll_{c,q} \frac{1}{\|\beta\|}\, q^{\nu(c-1)} + 1. \end{aligned} \tag{6.9}$$

Proof. Let S denote the considered sum. Using (5.4), we obtain

$$|S| \leqslant \left| \sum_{q^{(\nu-1)c} < m \leqslant x^c} \gamma m^{\gamma-1}\, \mathrm{e}(\beta m) \right| + \frac{1}{4}.$$

Partial summation yields

$$\begin{aligned}\sum_{q^{(\nu-1)c} < m \leqslant x^c} \gamma m^{\gamma-1}\, \mathrm{e}(\beta m) &= \gamma x^{c(\gamma-1)} \sum_{q^{(\nu-1)c} < m \leqslant x^c} \mathrm{e}(\beta m) \\ &\quad - \gamma(\gamma-1) \int_{q^{(\nu-1)c}}^{x^c} \sum_{q^{(\nu-1)c} < m \leqslant u} \mathrm{e}(\beta m)\, u^{\gamma-2}\, \mathrm{d}u. \end{aligned}$$

Since $\beta \notin \mathbb{Z}$, we have for all $q^{(\nu-1)c} < u \leqslant x^c$ that

$$\left| \sum_{q^{(\nu-1)c} < m \leqslant u} \mathrm{e}(\beta m) \right| \leqslant \frac{1}{|\sin \pi\beta|}.$$

We get (note, that $x \leqslant q^\nu$)

$$\begin{aligned} S &\leqslant \frac{\gamma}{|\sin \pi\beta|}\left(q^{(\nu-1)c(\gamma-1)} - \int_{q^{(\nu-1)c}}^{x^c} (\gamma-1) u^{\gamma-2}\, \mathrm{d}u \right) + \frac{1}{4} \\ &\leqslant \frac{\gamma}{|\sin \pi\beta|}\, q^{(\nu-1)(1-c)}\left(2 - q^{1-c}\right) + \frac{1}{4}, \end{aligned}$$

and the result follows. □

Lemma 6.10. *Let $c \in (1,2)$ and $\beta \in \mathbb{R}$. Furthermore, let x and ν be integers with $q^{\nu-1} < x \leqslant q^\nu$. Then we have*

$$\sum_{q^{(\nu-1)c} < m \leqslant x^c} \mathrm{e}(\beta m)\left(\Psi(-(m+1)^\gamma) - \Psi(-m^\gamma)\right) \ll_q \nu q^{\nu(1-(2-c)/3)}.$$

Proof. We can write

$$\begin{aligned}
&\sum_{q^{(\nu-1)c}<n\leqslant x^c} \mathrm{e}\left(\beta m\right)\left(\Psi(-(m+1)^\gamma)-\Psi(-m^\gamma)\right) \\
&\quad= \sum_{0\leqslant j<c\frac{\log q}{\log 2}} \sum_{\substack{q^{(\nu-1)c}2^j<n\leqslant q^{(\nu-1)c}2^{j+1}\\ q^{(\nu-1)c}<n\leqslant x}} \mathrm{e}\left(\beta m\right)\left(\Psi(-(m+1)^\gamma)-\Psi(-m^\gamma)\right) \\
&\quad\ll_q \max_{q^{(\nu-1)c}\leqslant M\leqslant q^{\nu c}} \max_{M<M'\leqslant 2M} \sum_{M<n\leqslant M'} \mathrm{e}\left(\beta m\right)\left(\Psi(-(m+1)^\gamma)-\Psi(-m^\gamma)\right).
\end{aligned}$$

In order to prove Lemma 6.10, it suffices to show that for $M > q^{(\nu-1)c}$ we have

$$\begin{aligned}
S_M := &\left|\sum_{M<m\leqslant M'} \mathrm{e}\left(\beta m\right)\left(\Psi(-(m+1)^\gamma)-\Psi(-m^\gamma)\right)\right| \\
&\ll (\log M)M^{\gamma(1-(2-c)/3)}. \qquad (6.10)
\end{aligned}$$

Approximating the function Ψ by trigonometric polynomials, we obtain exactly the same way as in Section 5.3.1 that

$$\begin{aligned}
S_M \leqslant &\left|\sum_{M<m\leqslant M'} \mathrm{e}\left(\beta m\right)\left(\Psi_{K_0}(-(m+1)^\gamma)-\Psi_{K_0}(-m^\gamma)\right)\right| \\
&\qquad + \frac{1}{2K_0+2}\sum_{M<m\leqslant M'} \kappa_{K_0}\left(-(m+1)^\gamma\right) + \frac{1}{2K_0+2}\sum_{M<m\leqslant M'} \kappa_{K_0}\left(-m^\gamma\right),
\end{aligned}$$

where $K_0 := \lfloor M^{1-\gamma(1-\delta)} \rfloor$ with a constant $\delta > 0$ that we choose later on. The polynomials $\kappa_K(t)$ and Ψ_{K_0} are given by (5.7), where $\kappa_K(t)$ is the periodic and positive Fejer kernel (see also Proposition A.20). Recall that

$$\frac{1}{2K_0+2}\sum_{M\leqslant m\leqslant 2M} \kappa_K(m^\theta) \ll_\theta K^{-1}M + K^{1/2}M^{\theta/2} + K^{-1/2}M^{1-\theta/2}$$

for every $0 < \theta < 1$ and for every $M \geqslant 1$ (see (5.8)). We obtain

$$\begin{aligned}
S_M \leqslant &\left|\sum_{M<m\leqslant M'} \mathrm{e}\left(\beta m\right)\left(\Psi_{K_0}(-(m+1)^\gamma)-\Psi_{K_0}(-m^\gamma)\right)\right| \\
&\qquad + K_0^{-1}M + K_0^{1/2}M^{\gamma/2} + K_0^{-1/2}M^{1-\gamma/2}.
\end{aligned}$$

For our special choice of K_0 we have that

$$K_0^{1/2}M^{\gamma/2} = M^{1/2+\gamma\delta/2} \geqslant M^{1/2-\gamma\delta/2} = K_0^{-1/2}M^{1-\gamma/2}.$$

Thus we get

$$\begin{aligned}
S_M \ll &\left|\sum_{M<m\leqslant M'} \mathrm{e}\left(\beta m\right)\left(\Psi_{K_0}(-(m+1)^\gamma)-\Psi_{K_0}(-m^\gamma)\right)\right| \\
&+ M^{\gamma(1-\delta)} + M^{1/2+\gamma\delta/2}. \qquad (6.11)
\end{aligned}$$

Next we treat the sum that arises in (6.11). Replacing Ψ_{K_0} by its expression and following exactly the same steps as in [MR05, Section 2.3] (we omit the details, since these steps are also similar to the ones in Section 5.3.1), we obtain

$$\begin{aligned}\sum_{M<m\leqslant M'} &\mathrm{e}\left(\beta m\right)\left(\Psi_{K_0}(-(m+1)^\gamma) - \Psi_{K_0}(-m^\gamma)\right) \\ &\ll (\log K_0) \max_{0<K\leqslant K_0} \max_{u\in\{0,1\}} \max_{\tilde{M}\in[M,2M]} \min\left(M^{1-\gamma}, K^{-1}\right) S(K,\tilde{M},u), \qquad (6.12)\end{aligned}$$

where $S(K,\tilde{M},u)$ is defined by (6.8). In the considered interval $[M,2M]$ we have the estimate

$$|k|M^{\gamma-2} \ll \left|\frac{\mathrm{d}^2\left(\beta y + k(y+u)^\gamma\right)}{\mathrm{d}y^2}\right| \ll |k|M^{\gamma-2}.$$

It follows from Theorem A.13 that

$$\begin{aligned}S(K,\tilde{M},u) &\ll \sum_{K<k\leqslant 2K}\left(k^{1/2}M^{\gamma/2} + k^{-1/2}M^{1-\gamma/2}\right) \\ &\ll K^{3/2}M^{\gamma/2} + K^{1/2}M^{1-\gamma/2}.\end{aligned}$$

If $K \leqslant M^{1-\gamma}$ we have

$$M^{\gamma-1}S(K,\tilde{M},u) \ll K^{3/2}M^{3\gamma/2-1} + K^{1/2}M^{\gamma/2} \ll M^{1/2},$$

whereas

$$K^{-1}S(K,\tilde{M},u) \ll K^{1/2}M^{\gamma/2} + K^{-1/2}M^{1-\gamma/2} \ll K^{1/2}M^{\gamma/2} + M^{1/2}$$

if $K > M^{1-\gamma}$. With (6.12) and the definition of K_0 we get

$$\begin{aligned}\sum_{M<m\leqslant M'} &\mathrm{e}\left(\beta m\right)\left(\Psi_{K_0}(-(m+1)^\gamma) - \Psi_{K_0}(-m^\gamma)\right) \\ &\ll (\log K_0)\left(K_0^{1/2}M^{\gamma/2} + M^{1/2}\right) \ll (\log M)M^{1/2+\gamma\delta/2}.\end{aligned}$$

Finally, we have (see (6.11))

$$S_M \ll (\log M)\left(M^{\gamma(1-\delta)} + M^{1/2+\gamma\delta/2}\right).$$

Now we choose $\delta > 0$ such that the upper bound is as small as possible. This is apparently the case if $\delta = (2-c)/3$ and we are done. □

Proof of Proposition 6.8. The Proposition now follows immediately from equation (6.7) and the previous two lemmas. □

6.3 Exponential sums for $c > 1$, $c \notin \mathbb{N}$

In this section we give a nontrivial upper bound of the sum $\sum_n \mathrm{e}\left(\beta \left\lfloor n^c \right\rfloor\right)$ for all real numbers $c > 1$ which are different from an integer. If $1 < c < 19/11$, then it turns out that the method based on Mauduit's and Rivat's work gives a better result (see Remark 6.13).

If $\|\beta\|$ is relatively small, then the estimation of $\sum_n \mathrm{e}\left(\beta \left\lfloor n^c \right\rfloor\right)$ can be reduced to a similar problem where $\mathrm{e}\left(\beta \lfloor n^c \rfloor\right)$ is replaced by $\mathrm{e}\left(\beta n^c\right)$. This leads to a simple application of the Kusmin-Landau Theorem (Theorem A.12). In the other case, we enhance a method of Deshouillers to obtain a nontrivial upper bound.

Proposition 6.11. *Let c be a real number > 1 and x and ν be integers such that $q^{\nu-1} < x \leqslant q^\nu$. Furthermore, let $\beta \in \mathbb{R}$ with $0 < \|\beta\| < \frac{1}{2c}\, q^{\nu(1-c)}$. Then we have*

$$\sum_{q^{\nu-1} < n \leqslant x} \mathrm{e}\left(\beta \left\lfloor n^c \right\rfloor\right) \ll_{c,q} \frac{1}{\|\beta\|}\, q^{\nu(1-c)} + q^{\nu(2-c)}. \tag{6.13}$$

Proof. Let S be the sum considered in (6.13). Without loss of generality, we can assume that $0 < \beta < \frac{1}{2c}\, q^{\nu(1-c)}$. Since

$$\mathrm{e}(\beta \left\lfloor n^c \right\rfloor) = \mathrm{e}\left(\beta n^c\right) \mathrm{e}\left(-\beta\{n^c\}\right) = \mathrm{e}(\beta n^c)\left(1 + O\left(\beta\right)\right),$$

we obtain

$$|S| = \left| \sum_{q^{\nu-1} < n \leqslant x} \mathrm{e}(\beta n^c)\, \mathrm{e}(-\beta\{n^c\}) \right| \ll \left| \sum_{q^{\nu-1} < n \leqslant x} \mathrm{e}(\beta n^c) \right| + \frac{1}{2c}\, q^{\nu(2-c)}.$$

Thus, it suffices to consider the sum $\sum_{q^{\nu-1} < n \leqslant x} \mathrm{e}(\beta n^c)$. If we set $f(y) = \beta y^c$, then we have for $y \in [q^{\nu-1}, q^\nu]$ the estimate

$$c\beta q^{(\nu-1)(c-1)} \leqslant \left|f'(y)\right| \leqslant c\beta q^{\nu(c-1)} \leqslant 1/2.$$

Furthermore, $f''(y) \neq 0$ on the considered interval. Hence, we can use Theorem A.12 and get

$$\sum_{q^{\nu-1} < n \leqslant x} \mathrm{e}(\beta n^c) \ll_{c,q} \frac{1}{\beta} q^{\nu(1-c)}.$$

This proves the desired result. □

In order to state the next proposition, we define the constant $\rho = \rho(c)$ by

$$\rho := \max\left(\rho_1, \rho_2, \rho_3, \rho_4\right), \tag{6.14}$$

where $\rho_1 = \frac{\lfloor c \rfloor + 1 - c}{2^{\lfloor c \rfloor + 1} - 1}$, $\rho_2 = \frac{\lfloor c \rfloor + 2 - c}{2^{\lfloor c \rfloor + 2} - 1}$, $\rho_3 = \left(3 \left\lfloor c + \frac{301}{300} \right\rfloor^2 \log\left(125 \left\lfloor c + \frac{301}{300} \right\rfloor\right)\right)^{-1}$ and $\rho_4 = 2^{-18}\left(c + \frac{1}{2^{18} c^2}\right)^{-2}$.

See Figure 6.1 for the considered terms in the definition of ρ in the interval $[1, 4]$ and Figure 6.2 in the interval $[9, 12]$. If $c < 12 - 1365/(121 \log 1375) \approx 10.4388$, then ρ_1 and ρ_2 contribute to the size of ρ. If $c > 12 - 1365/(121 \log 1375)$ then $\rho = \rho_3$ until ρ_4 is significant.

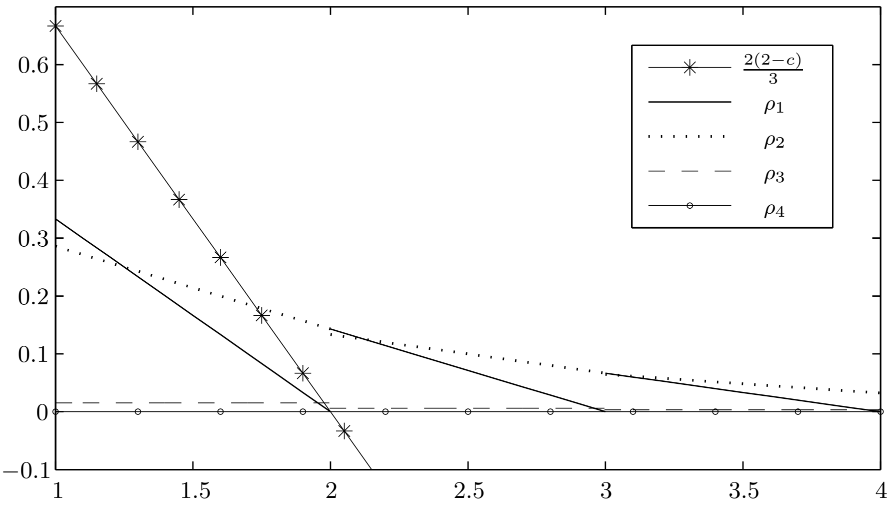


Figure 6.1: ρ in the interval $[1, 4]$

Proposition 6.12. *Let $c > 1$ be a real number different from an integer. Furthermore, let x and ν be integers with $q^{\nu-1} < x \leqslant q^\nu$ and $\beta \in \mathbb{R}$ such that $\|\beta\| \geqslant \frac{1}{2c} q^{\nu(1-c)}$. Then we have*

$$\sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}\left(\beta \lfloor n^c \rfloor\right) \ll_{c,q} \nu q^{\nu(1-\rho/2)}, \tag{6.15}$$

where ρ is defined by (6.14).

Remark 6.13. If $1 < c < 19/11$, then Proposition 6.8 implies Proposition 6.12. Indeed, $2(2-c)/3$ is greater than ρ in this case (see Figure 6.1) and the method of Mauduit and Rivat gives a better upper bound.

Remark 6.14. Let $c > 1$ be a real number different from an integer and x and ν be integers with $q^{\nu-1} < x \leqslant q^\nu$. If we set $\tilde{\rho} := \max(2(2-c)/3, \rho)$, then Proposition 6.8 implies together with Propositions 6.11 and 6.12

$$\sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}\left(\beta \lfloor n^c \rfloor\right) \ll_{c,q} \nu q^{\nu(1-\tilde{\rho}/2)} + \frac{1}{\|\beta\|} q^{\nu(1-c)}$$

for every $\beta \in \mathbb{R} \setminus \mathbb{Z}$.

As already pointed out, the method given in this section goes back to Deshouillers [Des73a]. He showed that if $c > 12$ ($c \notin \mathbb{N}$) and $\|\beta\|$ is not too small, then the considered sum (6.15) is of order $O(x^{1-\rho})$, where $\rho = (6c^2(\log c + 14))^{-1}$. We improve this result by enhancing two main tools of his method. On the one hand, we use van der Corput's method on exponential sums for small c and a refined version

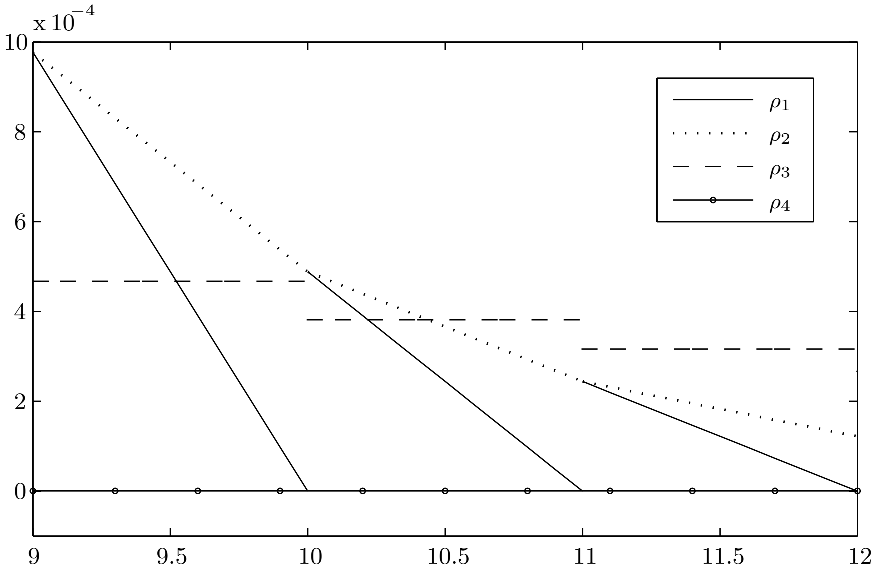


Figure 6.2: ρ in the interval $[9, 12]$

of Vinogradov's method on exponential sums for big c (see Lemma 6.15). On the other hand, we use the approximation properties of the Beurling-Selberg function summarized in Appendix A.5.

The method presented in this section cannot be applied for $c \in \mathbb{N}$. Note that Lemma 6.15 is false for integer exponents (take for example $\xi = 1$). The main difference for $c \in \mathbb{N}$ is that the m-th derivative of x^c is zero if $m \geqslant c + 1$ (cf. (6.17)). This makes it impossible to use van der Corput's and Vinogradov's method on exponential sums (even for $\xi < 1$).

Lemma 6.15. *Let $c > 1$ be a real number different from an integer and ρ defined by* (6.14)*. Furthermore, let x and ν be integers satisfying $q^{\nu-1} < x \leqslant q^{\nu}$ and let $\xi \in \mathbb{R}$ such that $\frac{1}{2c} q^{\nu(1-c)} \leqslant |\xi| \leqslant q^{(\nu-1)\rho}$. Then we have*

$$\sum_{q^{\nu-1} < n \leqslant x} \mathrm{e}(\xi n^c) \ll_{c,q} q^{\nu(1-\rho)}.$$

Proof. We can write

$$\sum_{q^{\nu-1} < n \leqslant x} \mathrm{e}(\xi n^c) = \sum_{0 \leqslant j < \frac{\log q}{\log 2}} \sum_{\substack{q^{\nu-1} 2^j < n \leqslant q^{\nu-1} 2^{j+1} \\ q^{\nu-1} < n \leqslant x}} \mathrm{e}(\xi n^c)$$
$$\ll_q \max_{q^{\nu-1} \leqslant M \leqslant q^{\nu}} \max_{M < M' \leqslant 2M} \sum_{M < n \leqslant M'} \mathrm{e}(\xi n^c).$$

Since we have for any $q^{\nu-1} \leqslant M \leqslant q^{\nu}$ that

$$\frac{1}{2c} q^{1-c} M^{1-c} \leqslant \frac{1}{2c} q^{\nu(1-c)} \leqslant |\xi| \leqslant q^{(\nu-1)\rho} \leqslant M^{\rho},$$

it suffices to show that for $M \geqslant 1$, $M < M' \leqslant 2M$ and $\frac{1}{2c}\, q^{1-c} M^{1-c} \leqslant |\xi| \leqslant M^{\rho}$ we have

$$\sum_{M<n\leqslant M'} \mathrm{e}(\xi n^c) \ll_{c,q} M^{1-\rho}. \tag{6.16}$$

We set $f(y) = \xi y^s$. Then we derive for every $m \geqslant 1$,

$$\left| \frac{y^m}{m!} f^{(m)}(y) \right| = |\xi| \left| \binom{c}{m} \right| y^c.$$

A short calculation shows that

$$\frac{\|c\|}{2m^{c+1}} \leqslant \left| \binom{c}{m} \right| \leqslant c^m.$$

Hence, there exists a constant $A = A(c, q) > 1$, such that

$$A^{-m} F \leqslant \left| \frac{y^m}{m!} f^{(m)}(y) \right| \leqslant A^m F \tag{6.17}$$

for every $y \in [M, 2M]$ and $m \geqslant 1$, where $F = |\xi| M^c$. In order to get a manageable notation, we set $\ell = (\log |\xi|)/(\log M)$. Then we have

$$M \ll \frac{1}{2c}\, q^{1-c} M^{1-c} M^c \leqslant |\xi| M^c = F = M^{\ell+c} \leqslant M^{\rho+c}.$$

We can apply Theorem A.14 (a van der Corput estimate) and obtain that for every $r \geqslant 0$,

$$\sum_{M<n\leqslant M'} \mathrm{e}(\xi n^c) \ll_{c,q,r} F^{\frac{1}{2^{r+2}-2}} M^{1-\frac{r+2}{2^{r+2}-2}} = M^{1-\frac{r+2-\ell-c}{2^{r+2}-2}}. \tag{6.18}$$

Let us fix c. Then we have that ρ is equal to one of the four possible choices ρ_1, ρ_2, ρ_3 or ρ_4 (see (6.14)). Recall that ρ can be equal to ρ_3 or ρ_4 only if $c \geqslant 12 - 1365/(121 \log 1375)$.

First, we assume that $\rho = \rho_1 = (\lfloor c \rfloor + 1 - c)/(2^{\lfloor c \rfloor + 1} - 1)$. Using inequality (6.18) with $r = \lfloor c \rfloor - 1$, we obtain

$$\sum_{M<n\leqslant M'} \mathrm{e}(\xi n^c) \ll_{c,q} M^{1-\frac{\lfloor c \rfloor + 1 - \ell - c}{2^{\lfloor c \rfloor + 1} - 2}} \ll_{c,q} M^{1-\rho_1}.$$

The last inequality follows from the fact that

$$\frac{\lfloor c \rfloor + 1 - \ell - c}{2^{\lfloor c \rfloor + 1} - 2} \geqslant \frac{\lfloor c \rfloor + 1 - \rho_1 - c}{2^{\lfloor c \rfloor + 1} - 2} = \rho_1. \tag{6.19}$$

Next we consider the case $\rho = \rho_2 = (\lfloor c \rfloor + 2 - c)/(2^{\lfloor c \rfloor + 2} - 1)$. We apply inequality (6.18) with $r = \lfloor c \rfloor$ and obtain

$$\sum_{M<n\leqslant M'} \mathrm{e}(\xi n^c) \ll_{c,q} M^{1-\frac{\lfloor c \rfloor + 2 - \ell - c}{2^{\lfloor c \rfloor + 2} - 2}} \ll_{c,q} M^{1-\rho_2}.$$

The same calculation as above (see (6.19)) verifies the last inequality. Note, that we cannot improve these estimates by employing (6.18) with other values of r. Indeed, it is easy to show that for $c > 1$,

$$\sup_{r \geqslant 0} \left(\frac{r+2-c}{2^{r+2}-1} \right) = \max \left(\frac{\lfloor c \rfloor + 1 - c}{2^{\lfloor c \rfloor + 1} - 1}, \frac{\lfloor c \rfloor + 2 - c}{2^{\lfloor c \rfloor + 2} - 1} \right).$$

If c is big (and ρ is small), then we use van der Corput's method in combination with Vinogradov's method. Let us assume that $\rho = \rho_3$. As already noticed, c must be bigger than 10 in this case. For $\ell < 10 - c$ we use (6.18) with $r = \lfloor c + \ell \rfloor$ and obtain

$$\sum_{M < n \leqslant M'} \mathrm{e}(\xi n^c) \ll_{c,q} M^{1 - \frac{1}{2^{\lfloor c + \ell \rfloor + 2} - 2}}.$$

Note, that $\lfloor c + \ell \rfloor \leqslant 9$ and that we have for $c > 10$,

$$\frac{1}{2^{11} - 2} > 0,000488 > 0,000382 > \frac{1}{3 \left\lfloor 10 + \frac{301}{300} \right\rfloor^2 \log \left(125 \left\lfloor 10 + \frac{301}{300} \right\rfloor \right)} \geqslant \rho_3.$$

Hence, we get

$$\sum_{M < n \leqslant M'} \mathrm{e}(\xi n^c) \ll_{c,q} M^{1 - \rho_3}.$$

If $10 - c \leqslant \ell \leqslant \rho$, then we have

$$M \leqslant M^{-\ell - c + \lfloor \ell + c + 1 \rfloor + 1} = F^{-1} M^{\lfloor \ell + c + 1 \rfloor + 1} \leqslant M^2,$$

and $\lfloor \ell + c + 1 \rfloor \geqslant 11$. This allows us to use Theorem A.15 (a Vinogradov type estimate). We get

$$\sum_{M < n \leqslant M'} \mathrm{e}(\xi n^c) \ll_{c,q} M^{1 - \frac{1}{3 \lfloor \ell + c + 1 \rfloor^2 \log(125 \lfloor \ell + c + 1 \rfloor)}} \ll_{c,q} M^{1 - \rho_3}$$

in this case too. It remains to consider the case $\rho = \rho_4 = 2^{-18} \left(c + \frac{1}{2^{18} c^2} \right)^{-2}$. Again, this is only possible if $c > 10 > 4$ and we employ (6.18) if $\ell < 4 - c$ with $r = \lfloor \ell + c \rfloor$. We get

$$\sum_{M < n \leqslant M'} \mathrm{e}(\xi n^c) \ll_{c,q} M^{1 - \frac{1}{2^{\lfloor c + \ell \rfloor + 2} - 2}} \ll_{c,q} M^{1 - \frac{1}{2^5 - 2}} \ll_{c,q} M^{1 - \rho_4}.$$

On the contrary, if $4 - c \leqslant \ell \leqslant \rho_4$, then we can write

$$M^4 = M^{4-c} M^c \leqslant M^{\ell + c} = F \leqslant M^{\rho_4 + c}.$$

Using this fact and (6.17), we can employ Theorem A.16 (again a Vinogradov type estimate) and obtain

$$\sum_{M < n \leqslant M'} \mathrm{e}(\xi n^c) \ll_{c,q} M^{1 - \frac{1}{2^{18} (\ell + c)^2}} \ll_{c,q} M^{1 - \frac{1}{2^{18} (\rho_4 + c)^2}} \ll_{c,q} M^{1 - \rho_4}.$$

This finally shows (6.16) and finishes the proof of Lemma 6.15. □

Proof of Proposition 6.12. We can assume that $\frac{1}{2c}\, q^{\nu(1-c)} \leqslant \beta \leqslant 1/2$. Let k be a positive integer (which we choose later) and set

$$I_\ell := \left[\frac{\ell}{k}, \frac{\ell+1}{k}\right) \qquad \ell = 0, \ldots, k-1.$$

We start with the following correlation:

$$\sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}\left(\beta \lfloor n^c \rfloor\right) = \sum_{0\leqslant \ell<k} \sum_{\substack{q^{\nu-1}<n\leqslant x\\ \{n^c\}\in I_\ell}} \mathrm{e}(\beta \lfloor n^c \rfloor).$$

If $\{n^c\} \in I_\ell$, then there exists a real number $0 \leqslant \theta < 1$, such that

$$\mathrm{e}(\beta \lfloor n^c \rfloor) = \mathrm{e}\left(\beta n^c - \beta\frac{\ell}{k} - \beta\frac{\theta}{k}\right) = \mathrm{e}\left(\beta n^c - \beta\frac{l}{k}\right)\left(1 + O\left(\frac{1}{k}\right)\right).$$

Thus, we obtain

$$\left|\sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}\left(\beta \lfloor n^c \rfloor\right)\right| \ll \sum_{0\leqslant \ell<k} \left|\sum_{\substack{q^{\nu-1}<n\leqslant x\\ \{n^c\}\in I_\ell}} \mathrm{e}(\beta n^c)\right| + \frac{q^\nu}{k}. \tag{6.20}$$

If we set $f_\ell(x) := \mathbf{1}_{I_\ell}(\{x\})$, where $\mathbf{1}_A$ denotes the characteristic function of the set A, then inequality (6.20) reads as follows:

$$\left|\sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}\left(\beta \lfloor n^c \rfloor\right)\right| \ll \sum_{0\leqslant \ell<k} \left|\sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}(\beta n^c) f_\ell(n^c)\right| + \frac{q^\nu}{k}. \tag{6.21}$$

Next, we approximate the function f_ℓ by trigonometric polynomials. Let $H \geqslant 1$ be an integer. Then there exist coefficients $a_H(h)$ with $|a_H(h)| \leqslant 2$, such that the trigonometric polynomial

$$f_{\ell,H}^*(t) = \frac{1}{k} + \frac{1}{2\pi i} \sum_{1\leqslant |h|\leqslant H} \frac{a_H(h)}{h}\, \mathrm{e}(ht)$$

verifies

$$|f_\ell(t) - f_{\ell,H}^*(t)| \leqslant \frac{1}{2H+2}\left(\kappa_H\left(t - \frac{\ell}{k}\right) + \kappa_H\left(t - \frac{\ell+1}{k}\right)\right),$$

where κ_H is the Fejer kernel defined by (5.7) (see Corollary A.22). We obtain (the integer H is chosen in the last step of the proof)

$$\left|\sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}(\beta n^c) f_\ell(n^c)\right| \leqslant \left|\sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}(\beta n^c) f_{\ell,H}^*(n^c)\right| + R(H), \tag{6.22}$$

where

$$R(H) := \frac{1}{2H+2} \sum_{q^{\nu-1}<n\leqslant x} \left(\kappa_H \left(n^c - \frac{\ell}{k} \right) + \kappa_H \left(n^c - \frac{\ell+1}{k} \right) \right).$$

The error term $R(H)$ can be estimated by

$$\frac{1}{2H+2} \sum_{q^{\nu-1}<n\leqslant x} \sum_{0\leqslant |h|\leqslant H} \left(1 - \frac{|h|}{H+1}\right) \left(1 + \mathrm{e}\left(-\frac{h}{k}\right)\right) \mathrm{e}\left(-\frac{h\ell}{k}\right) \mathrm{e}\left(hn^c\right)$$
$$\leqslant \frac{2}{2H+2} \sum_{0\leqslant |h|\leqslant H} \left| \sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}\left(hn^c\right) \right|.$$

We separate the case $h = 0$ from $h \neq 0$ and apply Lemma 6.15. This is admissible as long as $H \leqslant q^{(\nu-1)\rho}$, where ρ is defined by (6.14). We obtain

$$R(H) \ll_{c,q} \frac{q^\nu}{H} + q^{\nu(1-\rho)}.$$

Next, we use the definition of $f^*_{\ell,H}$ to deal with the first expression in the right-hand side of (6.22). We can write

$$\left| \sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}(\beta n^c) f^*_{\ell,H}(n^c) \right|$$
$$= \left| \sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}(\beta n^c) \left(\frac{1}{k} + \frac{1}{2\pi i} \sum_{1\leqslant |h|\leqslant H} \frac{a_H(h)}{h} \mathrm{e}(hn^c) \right) \right|$$
$$\leqslant \frac{1}{k} \left| \sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}(\beta n^c) \right| + \sum_{1\leqslant |h|\leqslant H} \frac{1}{h} \left| \sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}((\beta+h)n^c) \right|.$$

Applying Lemma 6.15 again (if $H \leqslant q^{(\nu-1)\rho}$), this is bounded by

$$\frac{q^{\nu(1-\rho)}}{k} + q^{\nu(1-\rho)} \log(H).$$

We obtain

$$\sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}(\beta n^c) f_\ell(n^c) \ll_{c,q} \frac{q^\nu}{H} + q^{\nu(1-\rho)} \log(H).$$

Together with inequality (6.21) this yields

$$\sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}\left(\beta \lfloor n^c \rfloor\right) \ll_{c,q} \frac{kq^\nu}{H} + kq^{\nu(1-\rho)} \log(H) + \frac{q^\nu}{k}.$$

If we set $k = \lfloor q^{(\nu\rho)/2} \rfloor$ and $H = \lfloor q^{(\nu-1)\rho} \rfloor$ (which actually verifies that we were allowed to use Lemma 6.15), we obtain

$$\sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}\left(\beta \lfloor n^c \rfloor\right) \ll_{c,q} \nu q^{\nu(1-\rho/2)}.$$

Finally, this implies the desired result. □

6.4 Proof of Theorem 6.2

In this section we prove Theorem 6.2. At first we shortly treat the (trivial) case $0 < c < 1$. The second part of the proof deals with the case $c > 1$ ($c \notin \mathbb{N}$) and it is based on methods coming from harmonic analysis (Appendix A.2) and on exponential sum estimates (Section 6.2 and 6.3).

6.4.1 Case: $0 < c < 1$

We set $\gamma = 1/c$ and $a_m := \#\{n \leqslant x : \lfloor n^c \rfloor = m\}$. Then we can write

$$\sum_{1 \leqslant n \leqslant x} \mathrm{e}\left(\alpha s_q(\lfloor n^c \rfloor)\right) = \sum_{1 \leqslant m \leqslant x^c} \mathrm{e}\left(\alpha s_q(m)\right) a_m.$$

If $m = \lfloor x^c \rfloor$, we have that $a_m = x - (\lfloor x^c \rfloor)^\gamma + O(1) = O(x^{1-c})$. If $m < \lfloor x^c \rfloor$, then we have $a_m = (m+1)^\gamma - m^\gamma + O(1) = \gamma m^{\gamma-1} + O(m^{\gamma-2} + 1)$. Since

$$\sum_{1 \leqslant m \leqslant x^c} \left(m^{\gamma-2} + 1\right) \ll_c x^{1-c} + x^c,$$

we obtain

$$\sum_{1 \leqslant n \leqslant x} \mathrm{e}\left(\alpha s_q(\lfloor n^c \rfloor)\right) \ll_c \sum_{1 \leqslant m \leqslant x^c} \mathrm{e}\left(\alpha s_q(m)\right) m^{\gamma-1} + x^{1-c} + x^c.$$

By partial summation we can write the occurring sum as

$$\begin{aligned}
&\sum_{1 \leqslant m \leqslant x^c} \mathrm{e}\left(\alpha s_q(m)\right) m^{\gamma-1} \\
&\qquad = x^{1-c} \sum_{1 \leqslant m \leqslant x^c} \mathrm{e}(\alpha s_q(m)) - (\gamma - 1) \int_1^{x^c} \sum_{1 \leqslant m < u} \mathrm{e}(\alpha s_q(m)) u^{\gamma-2} \, \mathrm{d}u.
\end{aligned}$$

Thus, we get

$$\sum_{1 \leqslant n \leqslant x} \mathrm{e}\left(\alpha s_q(\lfloor n^c \rfloor)\right) \ll_c x^{1-c} \max_{1 \leqslant N \leqslant x^c} \left| \sum_{1 \leqslant m \leqslant N} \mathrm{e}\left(\alpha s_q(m)\right) \right| + x^{1-c} + x^c. \tag{6.23}$$

A simple calculation shows that

$$\left| \sum_{0 \leqslant m < N} \mathrm{e}(\alpha s_q(m)) \right| \ll_q N^{\log_q \varphi_q(\alpha)}, \tag{6.24}$$

where $\varphi_q(t)$ is defined for all $t \in \mathbb{R}$ by

$$\varphi_q(t) = \begin{cases} \frac{|\sin \pi q t|}{|\sin \pi t|}, & \text{if } t \in \mathbb{R} \setminus \mathbb{Z}, \\ q, & \text{if } t \in \mathbb{Z}. \end{cases}$$

(See Appendix A.2.) Indeed, if $N = \sum_{j=0}^{\lambda} n_j q^j$ with $n_\lambda \neq 0$, then we can write

$$\sum_{0 \leqslant m < N} \mathrm{e}(\alpha s_q(m)) = \sum_{\nu=0}^{\lambda} \sum_{j_\nu=0}^{n_\nu - 1} \sum_{0 \leqslant m < q^\nu} \mathrm{e}\left(\alpha s_q\left(m + j_\nu q^\nu + n_{\nu+1} q^{\nu+1} + \cdots + n_\lambda q^\lambda\right)\right).$$

Since the sum-of-digits function is q-additive (that is, $s_q(a + bq^j) = s_q(a) + s_q(b)$ for $a < q^j$), we obtain

$$\left|\sum_{0 \leqslant m < N} \mathrm{e}(\alpha s_q(m))\right| \ll_q \sum_{\nu=0}^{\lambda} \left|\sum_{0 \leqslant m < q^\nu} \mathrm{e}(\alpha s_q(m))\right| = \sum_{\nu=0}^{\lambda} q^\nu \left|F_\nu(0,\alpha)\right| = \sum_{\nu=0}^{\lambda} \varphi_q(\alpha)^\nu,$$

where F_λ is defined by (6.6). Hence, we obtain the desired inequality (6.24). By local expansion (see for example Lemma A.6) we have

$$\varphi_q(t) \leqslant q^{1-\sigma_q' \|t\|^2},$$

where σ_q' is a positive computable constant only depending on q. Together with (6.23) and (6.24) this implies Theorem 6.2 for $0 < c < 1$.

6.4.2 Case: $c > 1$

For the following part we assume that x and ν are integers such that $q^{\nu-1} < x \leqslant q^\nu$. We set

$$S := \sum_{q^{\nu-1} < n \leqslant x} \mathrm{e}(\alpha s_q(\lfloor n^c \rfloor),$$

and use the abbreviation

$$\lambda := \lfloor \nu c \rfloor + 1. \tag{6.25}$$

Then we can write

$$\begin{aligned} S &= \sum_{0 \leqslant u < q^\lambda} \sum_{q^{\nu-1} < n \leqslant x} \mathrm{e}(\alpha s_q(u)) \cdot \frac{1}{q^\lambda} \sum_{0 \leqslant h < q^\lambda} \mathrm{e}\left(\frac{h(\lfloor n^c \rfloor - u)}{q^\lambda}\right) \\ &= \sum_{0 \leqslant h < q^\lambda} \frac{1}{q^\lambda} \sum_{0 \leqslant u < q^\lambda} \mathrm{e}\left(\alpha s_q(u) - huq^{-\lambda}\right) \sum_{q^{\nu-1} < n \leqslant x} \mathrm{e}\left(\frac{h\lfloor n^c \rfloor}{q^\lambda}\right). \end{aligned}$$

Using the Fourier transform defined in (6.6), we have

$$|S| \leqslant \sum_{0 \leqslant h < q^\lambda} |F_\lambda(h,\alpha)| \cdot \left|\sum_{q^{\nu-1} < n \leqslant x} \mathrm{e}\left(\frac{h\lfloor n^c \rfloor}{q^\lambda}\right)\right|. \tag{6.26}$$

It follows from Lemma A.7 that the contribution of the term, where $h = 0$, is bounded above by

$$|F_\lambda(0,\alpha)| q^\nu \ll q^{\nu - \sigma_q \|(q-1)\alpha\|^2 \lambda},$$

with a constant $c_q > 0$. If $0 < h < q^\lambda$, Remark 6.14 implies

$$\sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}\left(\frac{h\lfloor n^c\rfloor}{q^\lambda}\right) \ll_{c,q} \nu q^{\nu(1-\tilde{\rho}/2)} + \frac{q^\nu}{\min(h, q^\lambda - h)},$$

where $\tilde{\rho} := \max(2(2-c)/3, \rho)$. We obtain (using Lemma A.7 and A.8)

$$\sum_{0\leqslant h<q^\lambda} |F_\lambda(h,\alpha)| \left(\nu q^{\nu(1-\tilde{\rho}/2)} + \frac{q^\nu}{\min(h, q^\lambda - h)}\right)$$
$$\ll \nu q^{\nu(1-\tilde{\rho}/2)+\lambda\eta_q} + \log(q^\lambda) q^{\nu-\sigma_q\|(q-1)\alpha\|^2\lambda}.$$

Thus, we can bound the sum S by

$$S \ll_{c,q} \nu \left(q^{\nu(1-\sigma_q\|(q-1)\alpha\|^2 c)} + q^{\nu(1-\tilde{\rho}/2+c\eta_q)}\right).$$

If q is big enough (bigger than some constant $q_0(c)$), then it follows from Remark A.9 that

$$\tilde{\rho}/2 - c\eta_q > 0. \tag{6.27}$$

Setting $\sigma_{c,q} = \min(\sigma_q c, \tilde{\rho}/2 - \eta_q c) > 0$, we have for every $q \geqslant q_0(c)$ that

$$\sum_{q^{\nu-1}<n\leqslant x} \mathrm{e}(\alpha s_q(\lfloor n^c\rfloor) \ll_{c,q} \nu q^{\nu(1-\sigma_{c,q}\|(q-1)\alpha\|^2)}.$$

The proof of Theorem 6.2 is a direct consequence of this fact. Let ν_0 be the integer such that $q^{\nu_0-1} < x \leqslant q^{\nu_0}$. Then we can write

$$\sum_{1\leqslant n\leqslant x} \mathrm{e}\left(\alpha s_q(\lfloor n^c\rfloor)\right)$$
$$= \sum_{0\leqslant\nu<\nu_0} \sum_{q^{\nu-1}<n\leqslant q^\nu} \mathrm{e}\left(\alpha s_q(\lfloor n^c\rfloor)\right) + \sum_{q^{\nu_0-1}<n\leqslant x} \mathrm{e}\left(\alpha s_q(\lfloor n^c\rfloor)\right)$$
$$\ll_{c,q} \sum_{0\leqslant\nu\leqslant\nu_0} \nu q^{\nu(1-\sigma_{c,q}\|(q-1)\alpha\|^2)} \ll_{c,q} \nu_0 q^{\nu_0(1-\sigma_{c,q}\|(q-1)\alpha\|^2)}.$$

Since $\nu_0 \leqslant \lfloor \log x/\log q + 1\rfloor$, we obtain

$$\sum_{1\leqslant n\leqslant x} \mathrm{e}\left(\alpha s_q(\lfloor n^c\rfloor))\right) \ll_{c,q} (\log x) x^{1-\sigma_{c,q}\|(q-1)\alpha\|^2}.$$

Finally, note that we can see from equation (6.27) that the constant $\tilde{\rho}$ determines the size of an admissible (and computable) value $q_0(c)$. Equation (6.27) is satisfied if $\log\log q/\log q < \tilde{\rho}/(2c)$. This implies for example that such an admissible value is given by Kc^{c^4}, where K is an absolute constant.

6.5 Proofs of Corollary 6.4 and Corollary 6.6

In order to show Corollary 6.4 we need information on the distribution of $\lfloor n^c \rfloor$ in arithmetic progressions. For $1 < c < 2$ this has been studied for example in [Des73b] (see also [Rie67, War74]) and for $c > 12$ (not an integer) in [Des73b]. For the convenience of the reader we state and prove the following lemma which holds true for all non-integral reals $c > 1$. It confirms the already known result for $1 < c < 2$ and slightly improves the known results in the other cases. Note that a shorter proof can be obtained by using Proposition 6.12 directly. However, then the exponent $1-\rho$ in equation (6.28) has to be replaced by $1-\rho/2$.

Lemma 6.16. *Let $c > 1$ be a real number different from an integer and a and d be integers with $d \geqslant 1$. Then we have*

$$\#\{n \leqslant x : \lfloor n^c \rfloor \equiv a \bmod d\} = \frac{x}{d} + O_{c,d}\left((\log x)x^{1-\rho}\right), \tag{6.28}$$

where ρ is defined by (6.14).

Proof. We begin with the following observation: The integer n satisfies $\lfloor n^c \rfloor \equiv a$ mod d if and only if $a/d \leqslant \{n^c/d\} < (a+1)/d$. In order to prove this lemma, it suffices to show that the discrepancy D of (n^c/d), where n ranges from 1 to x, can be bounded above by $D \ll_{c,d} (\log x)x^{-\rho}$. We use the Erdős-Turán inequality (see Theorem A.19 and Inequality (A.25)) saying that

$$D \leqslant \frac{1}{H+1} + \sum_{h=1}^{H} \frac{1}{h} \left| \frac{1}{x} \sum_{1 \leqslant n \leqslant x} e\left(\frac{h}{d} n^c\right) \right|,$$

where the integer $H > 0$ can be chosen arbitrarily. Let ν_0 be the smallest positive integer such that $1/d \geqslant \frac{1}{2c} 2^{\nu_0(1-c)}$ and let λ be defined by $2^{\lambda-1} < x \leqslant 2^{\lambda}$. Lemma 6.15 implies

$$\begin{aligned}\left| \sum_{1 \leqslant n \leqslant x} \mathrm{e}\left(\frac{h}{d} n^c\right) \right| &\leqslant 2^{\nu_0 - 1} + \sum_{\nu_0 \leqslant \nu \leqslant \lambda} \sum_{\substack{2^{\nu-1} < n \leqslant 2^{\nu} \\ n \leqslant x}} \mathrm{e}\left(\frac{h}{d} n^c\right) \\ &\ll_{c,d} \sum_{\nu=\nu_0}^{\lambda} 2^{\nu(1-\rho)} \ll_{c,d} x^{1-\rho},\end{aligned}$$

where ρ is defined by (6.14). If we set $H := \lfloor 2^{(\lambda-1)\rho} \rfloor$, then the Erdős-Turán inequality yields

$$\begin{aligned}D &\ll 2^{(1-\lambda)\rho} + \frac{1}{x} \sum_{h=1}^{\lfloor 2^{(\lambda-1)\rho} \rfloor} \frac{1}{h} \left| \sum_{1 \leqslant n \leqslant x} e\left(\frac{h}{d} n^c\right) \right| \\ &\ll_{c,d} 2^{(1-\lambda)\rho} + \log(2^{(\lambda-1)\rho}) x^{-\rho} \ll_{c,d} (\log x) x^{-\rho}.\end{aligned}$$

As indicated above, this shows the desired result. □

Proof of Corollary 6.4. We can write

$$\# \{n \leqslant x : s_q (\lfloor n^c \rfloor) \equiv a \bmod m\} = \sum_{n \leqslant x} \frac{1}{m} \sum_{0 \leqslant \ell < m} \mathrm{e}\left(\ell \frac{s_q (\lfloor n^c \rfloor) - a}{m}\right).$$

Let us first consider the case $0 < c < 1$. The main term comes from $\ell = 0$ and equals x/m. Due to Theorem 6.2 there exists a constant $\sigma'_{c,q,\ell/m}$ for every $1 \leqslant \ell < m$, such that

$$\sum_{n \leqslant x} \mathrm{e}\left(\frac{\ell}{m} s_q (\lfloor n^c \rfloor)\right) \ll_{c,q} (\log x) x^{1-\sigma'_{c,q,\ell/m}}.$$

The result follows by setting $\sigma_{c,q,m} = \min_{1 \leqslant \ell < m}(\sigma'_{c,q,\ell/m})$. If $c > 1$, then we put $d = (m, q-1)$, $m' = m/d$, $J = \{km' : 0 \leqslant k < d\}$ and $J' = \{0, \ldots, m-1\} \setminus J = \{km' + r : 0 \leqslant k < d, 1 \leqslant r < m'\}$. For $\ell = km' \in J$ we have,

$$\mathrm{e}\left(\frac{\ell}{m} s_q (\lfloor n^c \rfloor)\right) = \mathrm{e}\left(\frac{k}{d} s_q (\lfloor n^c \rfloor)\right) = \mathrm{e}\left(\frac{k}{d} \lfloor n^c \rfloor\right).$$

Hence, applying Lemma 6.16 leads us to

$$\begin{aligned} \frac{1}{m} \sum_{\ell \in J} \sum_{n \leqslant x} \mathrm{e}\left(\ell \frac{s_q (\lfloor n^c \rfloor) - a}{m}\right) &= \frac{d}{m} \sum_{\substack{n \leqslant x \\ \lfloor n^c \rfloor \equiv a \bmod d}} 1 \\ &= \frac{x}{m} + O_{c,d}\left((\log x) x^{1-\rho}\right). \qquad (6.29) \end{aligned}$$

If $J' = \emptyset$, Lemma 6.16 already implies Corollary 6.4 (we can choose $\sigma_{c,q,m} = (9/10)\rho$). If $J' \neq \emptyset$, we set $q' = (q-1)/d$. Since $(q', m') = 1$, we obtain for $\ell = km' + r \in J'$,

$$\frac{(q-1)\ell}{m} = \frac{dq'(km' + r)}{dm'} = q'k + \frac{q'r}{m'} \notin \mathbb{Z}.$$

Theorem 6.2 implies that there exists a constant $\sigma'_{c,q,\ell/m}$ for every $\ell \in J'$, such that

$$\sum_{n \leqslant x} \mathrm{e}\left(\frac{\ell}{m} s_q (\lfloor n^c \rfloor)\right) \ll_{c,q,m} (\log x) x^{1-\sigma'_{c,q,\ell/m}}.$$

Put

$$\sigma_{q,m,c} = \frac{9}{10} \min\left(\min_{\ell \in J'}\left(\sigma'_{c,q,\ell/m}\right), \rho\right) > 0.$$

Together with (6.29) this proves Corollary 6.4. □

Proof of Corollary 6.6. If $\alpha \in \mathbb{Q}$, then the sequence $(\alpha s_q (\lfloor n^c \rfloor))_{n \in \mathbb{N}}$ takes modulo 1 only a finite number of values and is therefore not uniformly distributed modulo 1. If $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, then Theorem 6.2 and Weyl's criterion (Theorem A.18) imply the result. □

Chapter 7

Local results

It's a thing that non-mathematicians don't realize. Mathematics is actually an esthetic subject almost entirely.

*John Horton Conway (*1937)*

In this chapter we discuss local distribution results for the sum-of-digits function. In particular, we show that under certain assumptions on a function $g : \mathbb{N} \to \mathbb{N}$ we get an asymptotic expansion of the numbers $\#\{n \leqslant x : s_q(g(n)) = k\}$, where s_q denotes the ordinary sum-of-digits function in $\mathbb{N}$. The main idea in determining these numbers is the fact that we can write them in the form

$$\#\{n \leqslant x : s_q(g(n)) = k\} = \int_0^1 S(\alpha)\, \mathrm{e}(-\alpha k) \mathrm{d}\alpha,$$

where $S(\alpha)$ is given by $S(\alpha) = \sum_{n \leqslant x} \mathrm{e}(\alpha s_q(g(n)))$. Such exponential sums (for special functions $g(n)$) already appeared in previous chapters. In particular, we can use estimates that we have already proved in order to obtain a local result for the function $g(n) = \lfloor n^c \rfloor$. An analog problem is treated in the ring of Gaussian integers and we provide for example an asymptotic expansion of the numbers $\#\{z \in \mathcal{D}_N : s_q^{\mathbb{C}}(z^2) = k\}$ for arbitrary κ-$\mathbb{Z}[i]$ sequences $(\mathcal{D}_N)_{n \in \mathbb{N}}$ with $\kappa = 1/2$, where $s_q^{\mathbb{C}}$ denotes the complex sum-of-digits function in $\mathbb{Z}[i]$.

7.1 Introduction and main results

Let $q \geqslant 2$ and s_q be the sum-of-digits function in $\mathbb{N}$. We use $\mathrm{e}(x)$ for $e^{2\pi i x}$ and $\|x\|$ denotes the distance of the real number x to its nearest integer. It is well-known that the distribution of the sum-of-digits function can be approximated by a normal distribution. In particular we have

$$\frac{1}{x} \# \left\{ n < x : s_q(n) \leqslant \mu_q \log_q x + y \sqrt{\sigma_q^2 \log_q x} \right\} = \Phi(y) + o(1), \tag{7.1}$$

where μ_k and σ_k are defined for all $k \geqslant 2$ by[1]

$$\mu_k = \frac{k-1}{2}, \qquad \sigma_k^2 = \frac{k^2-1}{12} \tag{7.2}$$

and $\Phi(y)$ denotes the normal distribution function (see [KM68]). The sum-of-digits function of primes and polynomial subsequences also satisfy such a distribution result (see [Kát86] and [BK95]).

A local version of (7.1) can be found in [MS97] and in [FM05]. In the second cited work it is proved that

$$\#\{n < x : s_q(n) = \mu_q \lfloor \log_q n \rfloor + b(\lfloor \log_q n \rfloor)\} = \sqrt{\frac{6}{\pi(q^2-1)}} \frac{x}{\sqrt{\log_q x}} + O_K\left(\frac{x}{\log x}\right)$$

uniformly for any $x \geqslant 2$ and any $b : \mathbb{N} \to \mathbb{R}$ such that $|b(\nu)| \leqslant K v^{1/4}$ and $\mu_q \nu + b(\nu) \in \mathbb{N}$ for any $n \geqslant 1$.

Drmota, Mauduit, and Rivat recently enhanced a method developed by Bassily and Kátai [BK95] (see also [BK96]) in order to show the following result:

Theorem DMR (Drmota, Mauduit, and Rivat [DMR09]). *Let $q \geqslant 2$. We then have uniformly in $k > 0$ with $(k, q-1) = 1$*

$$\begin{aligned}\#\{p < x : p \text{ prim}, s_q(p) = k\} \\ = \frac{q-1}{\varphi(q-1)} \frac{\pi(x)}{\sqrt{2\pi\sigma_q^2 \log_q x}} \left(e^{-\frac{(k-\mu_q \log_q x)^2}{2\sigma_q^2 \log_q x}} + O_\varepsilon\left((\log x)^{-1/2+\varepsilon}\right)\right),\end{aligned}$$

where $\varepsilon > 0$ is arbitrary but fixed, $\varphi(\,\cdot\,)$ denotes Euler's totient function and $\pi(x)$ is equal to the number of primes less than or equal to x.

In this chapter we improve the error term given in this theorem and we generalize it to other subsequences than primes. Let $g : \mathbb{N} \to \mathbb{N}$ satisfy the following conditions:

(C1) There exist real numbers $0 < \gamma < \delta$, $\xi \geqslant 0$ and $g_k \geqslant 0$, $k = 0, \dots, q-1$ (only depending on g and q) such that $x^\gamma \ll g(x) \ll x^\delta$ for all $x \geqslant 2$, the integers $g(n)$ have less than or equal to $(\log_q g(x)) + \xi$ digits[2] in base q for all $n < x$ and

$$\frac{1}{x}\#\{n < x : g(n) \equiv k \bmod q-1\} = g_k + O_\varepsilon\left(\frac{(\log\log x)^{5+\varepsilon}}{\log x}\right)$$

for all $k \geqslant 0$ and all $\varepsilon > 0$.

(C2) There exist constants $c_1 > 0$ and $d_1 \geqslant 0$ such that

$$\sum_{n \leqslant x} \mathrm{e}\left(\alpha s_q\left(g(n)\right)\right) \ll x^{1-c_1\|(q-1)\alpha\|^2} (\log x)^{d_1}$$

uniformly for real α.

[1] We use this notation in the whole Chapter 7.

[2] That is, for every $n < x$ we can write $g(n)$ in the form $\sum_{j=0}^{\log_q g(x)+\xi} \varepsilon_j(g(n))q^j$, where $\varepsilon_j(g(n))$ denotes the unique j-th digit of $g(n)$ in base q.

(C3) Let $\kappa > 0$ and $\sigma > 1$. Then there exist constants $c_2 > 0$ and $d_2 \geqslant 0$ such that for all A and B with $(A, B) = 1$ we have

$$\sum_{n \leqslant x} \mathrm{e}\left(\frac{A}{B} g(n)\right) \ll x^{1-c_2(\log\log x)^\sigma/\log x}(\log x)^{d_2},$$

whenever $e^{\kappa(\log\log x)^\sigma} \ll B \ll g(x) e^{-\kappa(\log\log x)^\sigma}$.

Theorem 7.1. *Let $q \geqslant 2$ and $g : \mathbb{N} \to \mathbb{N}$ satisfy Conditions (C1)–(C3). Then we have uniformly for all $k \geqslant 0$*

$$\begin{aligned} &\frac{1}{x}\#\{n < x : s_q(g(n)) = k\} \\ &\qquad = \frac{g_k \cdot (q-1)}{\sqrt{2\pi\sigma_q^2 \log_q g(x)}} \cdot e^{-\Delta_k^2/2} + O_{g,q,\varepsilon}\left(\frac{(\log\log x)^{5+\varepsilon}}{\log x}\right), \end{aligned} \tag{7.3}$$

where $\Delta_k = \frac{k - \mu_q \log_q g(x)}{\sqrt{\sigma_q^2 \log_q g(x)}}$ and $\varepsilon > 0$ is arbitrary but fixed.

Remark 7.2. Conditions (C2) and (C3) are essential in the proof of Theorem 7.1 but Condition (C1) could be relaxed in different directions. However, we use Condition (C1) in this way since most of the examples we are interested in satisfy this condition. Furthermore, it simplifies the proof of Theorem 7.1 and make it more readable.

Next we treat local results of the complex sum-of-digits function. Let $q = -a \pm i$, $a \geqslant 1$ and set $Q = |q|^2$. In this chapter we denote the complex sum-of-digits function by $s_q^{\mathbb{C}}$ (we will suppress the superfix $\mathbb{C}$ if there is no danger of confusion with the ordinary sum-of-digits function).

First local results of $s_q^{\mathbb{C}}$ (and more generally of block additive functions) on the Gaussian integers were obtained by Drmota, Grabner and Liardet [DGL08]. They proved that if $|k - \mu_Q \log_Q N| \leqslant C\sqrt{\log_Q N}$ (for some $C > 0$) one has

$$\#\left\{|z|^2 < N : s_q^{\mathbb{C}}(z) = k\right\} = \frac{\pi N}{\sqrt{2\pi\sigma_Q^2 \log_Q N}}\left(e^{-\frac{(k-\mu_Q \log_Q N)^2}{2\sigma_Q^2 \log_Q N}} + O\left(\frac{1}{\sqrt{\log N}}\right)\right).$$

In what follows we prove a result in $\mathbb{Z}[i]$ analog to Theorem 7.1. Let $(\mathcal{C}_N)_{N\in\mathbb{N}}$ be a sequence of subsets of Gaussian integers with

- $\mathcal{C}_N \subseteq \mathcal{C}_{N+1}$,
- $\mathcal{C}_N \subseteq \{z \in \mathbb{Z}[i] : \max(|\Re(z)|, |\Im(z)|) \leqslant \sqrt{N}\}$, and
- there exists a constant $c > 0$, such that $cN \leqslant \#\mathcal{C}_N$.

Furthermore let $g : \mathbb{Z}[i] \to \mathbb{Z}[i]$ satisfy the following conditions:

(ℂ1) There exist real numbers $0 < \gamma < \delta$, $\xi \geqslant 0$ and $g_k \geqslant 0$, $k = 0, \dots, a^2 + 2a + 2$ (only depending on g and q) such that $|z|^\gamma \ll |g(z)| \ll |z|^\delta$ for all z with $|z| \geqslant 2$, the integers $g(z)$ have less than or equal to $(\log_Q |g(\lfloor\sqrt{N}\rfloor)|^2) + \xi$ digits[3] in base q for all $z \in \mathcal{C}_N$ and

$$\frac{1}{\#\mathcal{C}_N}\#\{z \in \mathcal{C}_N : g(z) \equiv k \bmod q-1\} = g_k + O\left(\frac{(\log\log N)^{11}}{\log N}\right)$$

for all $k \geqslant 0$.

(ℂ2) There exist constants $c_1 > 0$ and $d_1 \geqslant 0$ such that

$$\sum_{z \in \mathcal{C}_N} \mathrm{e}\left(\alpha s_q^{\mathbb{C}}\left(g(z)\right)\right) \ll N^{1-c_1\|(a^2+2a+2)\alpha\|^2}(\log N)^{d_1}$$

uniformly for real α.

(ℂ3) Let $\kappa > 0$ and $\sigma > 1$. There exist constants $c_2 > 0$ and $d_2 \geqslant 0$ such that for all A and B with $(A, B) = 1$ we have

$$\sum_{z \in \mathcal{C}_N} \mathrm{e}\left(\frac{A}{B}g(z)\right) \ll N^{1-c_2(\log\log N)^\sigma/\log N}(\log N)^{d_2},$$

whenever $e^{\kappa(\log\log N)^\sigma} \ll |B|^2 \ll |g(\lfloor\sqrt{N}\rfloor)|^2 e^{-\kappa(\log\log N)^\sigma}$.

Theorem 7.3. *Let $q = -a \pm i$, $a \geqslant 1$ and $(\mathcal{C}_N)_{N\in\mathbb{N}}$ be a sequence of subsets of Gaussian integers as stated above. Moreover, let $g : \mathbb{Z}[i] \to \mathbb{Z}[i]$ satisfy Conditions (ℂ1)-(ℂ3). Then we have uniformly for all integers $k \geqslant 0$,*

$$\begin{aligned}\frac{1}{\#\mathcal{C}_N}\#&\left\{z \in \mathcal{C}_N : s_q^{\mathbb{C}}(g(z)) = k\right\}\\ &= \frac{g_k \cdot (a^2+2a+2)}{\sqrt{2\pi\sigma_Q^2 \log_Q |g(\lfloor\sqrt{N}\rfloor)|^2}} \cdot e^{-\Delta_k^2/2} + O\left(\frac{(\log\log N)^{11}}{\log N}\right),\end{aligned}$$

where $\Delta_k = \frac{k - \mu_Q \log_Q |g(\lfloor\sqrt{N}\rfloor)|^2}{\sqrt{\sigma_Q^2 \log_Q |g(\lfloor\sqrt{N}\rfloor)|^2}}$.

Remark 7.4. One can obtain an error term in Theorem 7.3 similar to the one in Theorem 7.1. In order to improve the readability of the proof we just show the stated version.

In Section 7.2 we prove Theorem 7.1. Our considerations are inspired by the work of Drmota et al. [DMR09]. Section 7.3 deals with Theorem 7.3. Parts of the reasoning in the Gaussian integers are very similar to the real case but the approximation of the fundamental domain in ℂ is much more difficult than in ℝ (compare Lemma 7.6 and Lemma 7.13). In order to deal with this problem we use ideas worked out by Gittenberger and Thuswaldner [GT00]. In Section 7.4 we give some examples of functions g that satisfy the conditions stated above and we obtain corollaries from Theorem 7.1 and Theorem 7.3.

[3] We denote in the complex case the j-th digit of $g(z)$ as in the integers case by $\varepsilon_j(g(z))$.

7.2 The sum of digits in $\mathbb{N}$

Let us fix an integer $k \geqslant 0$ and set $I(x,k) = \{n < x : g(n) \equiv k \bmod q-1\}$. Since

$$\{n < x : s_q(g(n)) = k\} \subseteq I(x,k),$$

we see that (7.3) holds trivially true if $g_k = 0$. Thus we can assume that $g_k > 0$. In this case, Condition (C1) implies that it suffices to show

$$\#\{n < x : s_q(g(n)) = k\} = \frac{(q-1)R(x,k)}{\sqrt{2\pi\sigma_q^2 \log_q g(x)}} \left(e^{-\Delta_k^2/2} + O_{g,q,\varepsilon}\left(\frac{(\log\log x)^{5+\varepsilon}}{(\log x)^{1/2}} \right) \right),$$

where $R(x,k) = \#I(x,k)$. In what follows, the implied constants in the O-terms may depend on the function g and on the base q, and we omit the dependence in the notation. As already mentioned in the preface, we can write

$$\#\{n < x : s_q(g(n)) = k\} = \int_0^1 S(\alpha)\,\mathrm{e}(-\alpha k)d\alpha, \tag{7.4}$$

where

$$S(\alpha) := \sum_{0 \leqslant n < x} \mathrm{e}(\alpha s_q(g(n))).$$

We will split the integral in (7.4) up into two different domains. The main term of the integral arises from the domain where α is near to $\ell/(q-1)$, $\ell = 0, .., q-1$. To calculate this part, we use probabilistic methods in order to succeed (see Section 7.2.1). The remaining part can be treated with help of Condition (C2) and we give a proof of Theorem 7.1 in Section 7.2.2.

7.2.1 A probabilistic method

The following property, which can be interpreted as a generalization of a central limit theorem, is the main result of this section.

Proposition 7.5. *Let $q \geqslant 2$ and $\varepsilon > 0$. Then we have for every integer $k \geqslant 0$,*

$$\sum_{n \in I(x,k)} \mathrm{e}\left(\alpha s_q\left(g(n)\right)\right) = R(x,k)\,\mathrm{e}\left(\alpha\mu_q \log_q g(x)\right) \cdot \left(e^{-2\pi^2\alpha^2\sigma_q^2 \log_q g(x)} + O_\varepsilon\left(|\alpha|(\log\log x)^{3+\varepsilon}\right) \right) \tag{7.5}$$

uniformly for real α with $|\alpha| \leqslant (\log\log x)(\log x)^{-1/2}$.

Condition (C1) implies that there exists a constant ξ such that the integers $g(n)$ have less than or equal to

$$L := (\log_q g(x)) + \xi$$

digits for all $n < x$. In what follows, we will prove that

$$\sum_{n \in I(x,k)} \mathrm{e}\left(\alpha s_q\left(g(n)\right)\right) = R(x,k)\,\mathrm{e}\left(\alpha\mu_q L\right) \cdot \Big(e^{-2\pi^2\alpha^2\sigma_q^2 L} + O_\varepsilon\big(|\alpha|(\log L)^{3+\varepsilon}\big)\Big) \tag{7.6}$$

uniformly for real α with $|\alpha| \ll (\log L)L^{-1/2}$. This implies Proposition 7.5.

Equation (7.6) can be translated into a probabilistic language. If we assume that every number in the set $I(x,k)$ is equally likely, then the function which assigns each number its j-th digit is a random variable. Hence, the sum-of-digits function $S_x(n) := s_q(g(n))$ for $n < x$ also can be interpreted as a random variable. Using this model, formula (7.6) is equivalent to the relation (set $\alpha = t/(2\pi\sigma_q L^{1/2})$)

$$\varphi_1(t) := \mathbb{E}e^{it(S_x - L\mu_q)/(L\sigma_q^2)^{1/2}} = e^{-t^2/2} + O_\varepsilon\left(|t|\frac{(\log L)^{3+\varepsilon}}{L^{\frac{1}{2}}}\right) \tag{7.7}$$

that is uniform for $|t| \ll L^{1/2} \ll (\log\log x)$. Note, that $\varphi_1(t)$ is the characteristic function of $(S_x - L\mu_q)/(L\sigma_q^2)^{1/2}$.

In order to prove this, we approximate the sum-of-digits function with a sum of uniformly and independently distributed random variables (at the level of moments). Therefore we need some information on the joint distribution of their summands. If $\sigma > 1$ (and x is big enough), we set

$$L' = \#\left\{j \in \mathbb{Z} : (\log\log x)^\sigma \leqslant j \leqslant L - (\log\log x)^\sigma\right\} = L - 2(\log\log x)^\sigma + O(1).$$

Lemma 7.6. *Let $1 \leqslant d \leqslant L'$, $\sigma > 1$ and $0 < \varepsilon < \sigma$. Furthermore, let $j_1, j_2, \ldots, j_d$ and $\ell_1, \ell_2, \ldots, \ell_d$ be integers with*

$$(\log\log x)^\sigma \leqslant j_1 < j_2 < \cdots < j_d \leqslant L - (\log\log x)^\sigma$$

and $\ell_1, \ell_2, \ldots, \ell_d \in \{0, 1, \ldots, q-1\}$. Then there exists a constant $\tilde{c} > 0$ such that we have uniformly for $d \leqslant (\log\log x)^{\sigma-1}$

$$\frac{1}{R(x,k)}\,\#\left\{n \in I(x,k) : \varepsilon_{j_1}\left(g(n)\right) = \ell_1, \ldots, \varepsilon_{j_d}\left(g(n)\right) = \ell_d\right\} = q^{-d} + O_\sigma\left(e^{-\tilde{c}(\log\log x)^\sigma}\right).$$

Drmota et al. used Fourier theoretic tools in order to prove a similar result (see [DMR09, Lemma 4.5]). In what follows we will make use of the approximation properties of the Beurling-Selberg function (Vaaler's method), which are summarized in Appendix A.5.

Proof of Lemma 7.6. The j-th digit of m is equal to ℓ if and only if

$$\left\{\frac{m}{q^{\ell+1}}\right\} \in I_\ell := \left[\frac{\ell}{q}, \frac{\ell+1}{q}\right).$$

Let $f_\ell(x) := \mathbf{1}_{I_\ell}(\{x\})$, where $\mathbf{1}_A$ denotes the characteristic function of the set A. Then we have

$$\#\left\{n \in I(x,k) : \varepsilon_{j_1}(g(n)) = \ell_1, \dots, \varepsilon_{j_d}(g(n)) = \ell_d\right\} = \sum_{n\in I(x,k)} \prod_{r=1}^{d} f_{\ell_r}\left(\frac{g(n)}{q^{j_r+1}}\right).$$

Next, we approximate the function f_ℓ by trigonometric polynomials exactly the same way as in Chapter 6 (see the proof of Proposition 6.12). Let $H \geqslant 1$ be an integer. Then there exist coefficients $a_{\ell,H}(h)$ with $|a_{\ell,H}(h)| \leqslant 2$, such that the trigonometric polynomial

$$f^*_{\ell,H}(t) = \frac{1}{q} + \frac{1}{2\pi i} \sum_{1\leqslant |h| \leqslant H} \frac{a_{\ell,H}(h)}{h} \mathrm{e}(ht)$$

verifies

$$|f_\ell(t) - f^*_{\ell,H}(t)| \leqslant \frac{1}{2H+2}\left(\kappa_H\left(t - \frac{\ell}{q}\right) + \kappa_H\left(t - \frac{\ell+1}{q}\right)\right),$$

where κ_H is the Fejer kernel defined by (5.7) (see Corollary A.22). We obtain

$$\begin{aligned}\#&\left\{n \in I(x,k) : \varepsilon_{j_1}(g(n)) = \ell_1, \dots, \varepsilon_{j_d}(g(n)) = \ell_d\right\}\\ &= \sum_{n\in I(x,k)} \prod_{r=1}^{d} f^*_{\ell_r,H}\left(\frac{g(n)}{q^{j_r+1}}\right) + E(x,k),\end{aligned}$$

where

$$E(x,k) = \sum_{n\in I(x,k)} \prod_{r=1}^{d} f_{\ell_r}\left(\frac{g(n)}{q^{j_r+1}}\right) - \sum_{n\in I(x,k)} \prod_{r=1}^{d} f^*_{\ell_r,H}\left(\frac{g(n)}{q^{j_r+1}}\right).$$

The error term can be written as

$$\begin{aligned}E(x,k) = \sum_{n\in I(x,k)} \sum_{r=1}^{d} &\left(f_{\ell_r}\left(\frac{g(n)}{q^{j_r+1}}\right) - f^*_{\ell_r,H}\left(\frac{g(n)}{q^{j_r+1}}\right)\right)\\ &\cdot \prod_{r<u\leqslant d} f_{\ell_u}\left(\frac{g(n)}{q^{j_u+1}}\right) \prod_{1\leqslant v<r} f^*_{\ell_v,H}\left(\frac{g(n)}{q^{j_v+1}}\right).\end{aligned}$$

Using the trivial estimate $f_\ell \leqslant 1$ and $|f^*_\ell| \leqslant 3$, we obtain

$$E(x,k) \leqslant \max_{1\leqslant r\leqslant d} d\,3^d \sum_{n\in I(x,k)} \left| f_{\ell_r}\left(\frac{g(n)}{q^{j_r+1}}\right) - f^*_{\ell_r,H}\left(\frac{g(n)}{q^{j_r+1}}\right)\right|.$$

Hence we have

$$\begin{aligned}E(x,k) &\leqslant \max_{1\leqslant r\leqslant d} d\,3^d \sum_{n\in I(x,k)} \frac{1}{2H+2}\left(\kappa_H\left(\frac{g(n)}{q^{j_r+1}} - \frac{\ell}{k}\right) + \kappa_H\left(\frac{g(n)}{q^{j_r+1}} - \frac{\ell+1}{k}\right)\right)\\ &\leqslant \max_{1\leqslant r\leqslant d} \frac{2d\,3^d}{2H+2} \sum_{|h|\leqslant H} \left| \sum_{n\in I(x,k)} \mathrm{e}\left(\frac{h}{q^{j_r+1}} g(n)\right)\right|.\end{aligned}$$

Thus we have to treat the sum $\sum_{n\in I(x,k)} \mathrm{e}\left(\frac{h}{q^{j_r+1}}g(n)\right)$. If $h=0$ this sum equals $R(x,k)$. Otherwise we have

$$\sum_{n\in I(x,k)} \mathrm{e}\left(\frac{h}{q^{j_r+1}}g(n)\right) = \frac{1}{q-1}\sum_{s=0}^{q-2}\mathrm{e}\left(-\frac{sk}{q-1}\right)\sum_{n\leqslant x}\mathrm{e}\left(\left(\frac{h}{q^{j_r+1}}+\frac{s}{q-1}\right)g(n)\right). \tag{7.8}$$

Let A and B be defined in such a way that $\frac{h}{q^{j_r+1}}+\frac{s}{q-1}=\frac{A}{B}$ and $(A,B)=1$. We set

$$H=\left\lfloor e^{\frac{\log q}{2}(\log\log x)^\sigma}\right\rfloor, \tag{7.9}$$

and obtain

$$e^{\frac{\log q}{4}(\log\log x)^\sigma} \leqslant q^{j_r}/H \ll B \ll g(x)e^{-(\log q)(\log\log x)^\sigma}.$$

(See the assumptions of Lemma 7.6). Condition (C3) implies that for all $1\leqslant r\leqslant d$

$$\left|\sum_{n\in I(x,k)}\mathrm{e}\left(\left(\frac{h}{q^{j_r+1}}+\frac{s}{q-1}\right)g(n)\right)\right| \ll x^{1-c_2(\log\log x)^\sigma/\log x}(\log x)^{d_2}.$$

Since we have $x\ll R(x,k)$ and $d\ll(\log\log x)^{\sigma-\varepsilon}$, we get

$$\begin{aligned} E(x,k) &\ll d\,3^d R(x,k)\left(\frac{1}{H}+e^{-c_2(\log\log x)^\sigma}(\log x)^{d_2}\right)\\ &\ll R(x,k)e^{-\tilde{c}(\log\log x)^\sigma}, \end{aligned}$$

for an appropriate constant $\tilde{c}>0$. Now we calculate the main term. We have

$$\sum_{n\in I(x,k)}\prod_{r=1}^{d} f^*_{\ell_r,H}\left(\frac{g(n)}{q^{j_r+1}}\right) = \sum_{n\in I(x,k)}\prod_{r=1}^{d}\left(\frac{1}{q}+\frac{1}{2\pi i}\sum_{1\leqslant|h|\leqslant H}\frac{a_{\ell_r,H}(h)}{h}\mathrm{e}\left(\frac{h}{q^{j_r+1}}g(n)\right)\right).$$

This is the same as

$$\begin{aligned} \frac{R(x,k)}{q^d}+\sum_{r=1}^{d}\frac{1}{q^{d-r}(2\pi i)^r}\sum_{1\leqslant k_1<\ldots<k_r\leqslant d}\ \sum_{1\leqslant|h_1|\leqslant H}\frac{a_{\ell_{k_1},H}(h_1)}{h_1}\cdots\sum_{1\leqslant|h_r|\leqslant H}\frac{a_{\ell_{k_r},H}(h_r)}{h_r}\\ \cdot\sum_{n\in I(x,k)}\mathrm{e}\left(\left(\frac{h_1}{q^{j_{k_1}+1}}+\cdots+\frac{h_r}{q^{j_{k_r}+1}}\right)g(n)\right). \end{aligned}$$

If we can show that for $d\leqslant(\log\log x)^{\sigma-1}$

$$\sum_{n\in I(x,k)}\mathrm{e}\left(\left(\frac{h_1}{q^{j_{k_1}+1}}+\cdots+\frac{h_r}{q^{j_{k_r}+1}}\right)g(n)\right) \ll x^{1-c_2(\log\log x)^\sigma/\log x}(\log x)^{d_2} \tag{7.10}$$

independently of the chosen integers $r, k_1, \ldots, k_r, h_1, \ldots, h_r$, then we have (see (7.9))

$$\begin{aligned}\frac{1}{R(x,k)} \sum_{n \in I(x,k)} \prod_{r=1}^{d} f^*_{\ell_r,H}\left(\frac{g(n)}{q^{j_r+1}}\right) - \frac{1}{q^d}& \\ &\ll d(\log H)^d x^{1-c_2(\log\log x)^\sigma/\log x} (\log x)^{d_2} \\ &\ll d\left(\frac{\log q}{2}(\log\log x)^\sigma\right)^d e^{-c_2(\log\log x)^\sigma} (\log x)^{d_2}.\end{aligned}$$

This yields

$$\frac{1}{R(x,k)} \sum_{n \in I(x,k)} \prod_{r=1}^{d} f^*_{\ell_r,H}\left(\frac{g(n)}{q^{j_r+1}}\right) - \frac{1}{q^d} \ll e^{-\tilde{c}(\log\log x)^\sigma},$$

for an appropriate constant $\tilde{c} > 0$ (which we assume without loss of generality to be the same as above). Hence, it remains to show (7.10). If we write

$$\frac{h_1}{q^{j_{k_1}+1}} + \cdots + \frac{h_r}{q^{j_{k_r}+1}} + \frac{s}{q-1} = \frac{A'}{B'}$$

with $(A', B') = 1$ and $0 \leqslant s < q-1$, then we have

$$e^{\frac{\log q}{4}(\log\log x)^\sigma} \ll q^{j_{k_1}}/(dH) \ll B \ll g(x) e^{-(\log q)(\log\log x)^\sigma}.$$

Thus, Condition (C3) implies (7.10) (compare with (7.8)), which proves the desired result. □

In order to prove Proposition 7.5, we truncate the sum-of-digits function and approximate it appropriately. Let σ be a real number greater than 1 (which we choose at the end of the proof). Furthermore, let Z_j be a sequence of independent random variables with range $\{0, 1, \ldots, q-1\}$ and uniform probability distribution, and set

$$T_x := \sum_{(\log\log x)^\sigma \leqslant j \leqslant L-(\log\log x)^\sigma} \varepsilon_j(g(n)), \qquad \overline{T}_x := \sum_{(\log\log x)^\sigma \leqslant j \leqslant L-(\log\log x)^\sigma} Z_j.$$

Define the random variables X and Y by $X := (T_x - L'\mu_q)/(L'\sigma_q^2)^{1/2}$ and $Y := (\overline{T}_x - L'\mu_q)/(L'\sigma_q^2)^{1/2}$, and let $\varphi_2(t)$ be the characteristic function of X and $\varphi_3(t)$ the characteristic function of Y.

Lemma 7.7. *We have uniformly for all real t*

$$|\varphi_1(t) - \varphi_2(t)| \ll |t| \frac{(\log\log x)^\sigma}{L^{1/2}}.$$

Proof. From the definition of L, L', S_x and T_x it follows that $|L-L'| \ll (\log\log x)^\sigma$, $\|S_x - T_x\|_\infty \ll (\log\log x)^\sigma$ and that $\|T_x\|_\infty \ll L'$. Thus, we obtain (using the elementary estimate $|e^{it}-e^{is}| \leqslant |t-s|$)

$$\begin{aligned}|\varphi_1(t)-\varphi_2(t)| &\leqslant |t|\,\mathbb{E}\left|\frac{S_x - L\mu_q}{(L\sigma_q^2)^{1/2}} - \frac{T_x - L'\mu_q}{(L'\sigma_q^2)^{1/2}}\right| \\ &\ll |t|\mathbb{E}\left|\frac{S_x-T_x}{L^{1/2}} - \frac{(L-L')\mu_q}{L^{1/2}} + (T_x - L'\mu_q)\left(\frac{1}{L^{1/2}} - \frac{1}{L'^{1/2}}\right)\right| \\ &\ll |t|\left(\frac{\|S_x-T_x\|_\infty}{L^{1/2}} + \frac{|L-L'|}{L^{1/2}} + L'\left(\frac{1}{L'^{1/2}} - \frac{1}{L^{1/2}}\right)\right),\end{aligned}$$

and the result follows. □

Lemma 7.8. *We have*

$$\mathbb{E}e^{wY} = e^{w^2/2}\left(1 + O\left(\frac{|w|^4}{\log x}\right)\right)$$

uniformly for $|w| \leqslant (\log x)^{\frac{1}{4}}$.

Proof. See for example Lemma 4.2 of [DMR09], which is a slight variant of this statement (the proof is actually the same). □

In particular, the characteristic function of the normalized random variable $\overline{T}_x$ is given by

$$\varphi_3(t) = e^{-t^2/2}\left(1 + O\left(\frac{t^4}{L}\right)\right)$$

uniformly for $|t| \ll \log\log x$. We need two more auxiliary results before we can prove Proposition 7.5.

Lemma 7.9. *Let D be an even number with $D = o((\log x)^{1/2})$. Then we have*

$$\mathbb{E}\,Y^D \ll \frac{D!}{D^{D/2}\,e^{-D/2}\,D^{1/2}}.$$

Remark 7.10. If one carries out a more thorough analysis as in the proof given below, one can obtain an asymptotic expansion of $\mathbb{E}\,Y^D$ (see [DMR09, Section 4.3]).

Proof of Lemma 7.9. Lemma 7.8 and Cauchy's formula imply

$$\mathbb{E}\,Y^D = \frac{D!}{2\pi i}\int_{|w|=w_0} e^{w^2/2}\left(1 + O\left(\frac{w^4}{\log x}\right)\right)\frac{\mathrm{d}w}{w^{D+1}},$$

whenever w_0 is a positive real number satisfying $w_0 = o\big((\log x)^{1/2}\big)$. We (crudely) employ the saddle point method and choose $w_0 = D^{1/2}$. We get

$$\begin{aligned}
\mathbb{E}Y^D &= \frac{D!}{2\pi D^{D/2}} \int_{-\pi}^{\pi} e^{\frac{D}{2}e^{2i\varphi}-Di\varphi}\left(1+O\left(\frac{D^2}{L}\right)\right)\mathrm{d}\varphi \\
&\ll \frac{D!}{D^{D/2}} \int_0^{\pi/2} e^{\frac{D}{2}\cos(2\varphi)}\mathrm{d}\varphi \\
&\ll \frac{D!}{D^{D/2}} \int_0^{\pi/2} e^{\frac{D}{2}\left(1-2\varphi^2+O(\varphi^4)\right)}\mathrm{d}\varphi \\
&\ll \frac{D!}{D^{D/2}e^{-D/2}} \int_0^{\pi/2} e^{-D\varphi^2}\left(1+O(D\varphi^4)\right)\mathrm{d}\varphi.
\end{aligned}$$

Next we change variables ($\sqrt{2D}\varphi = \psi$) and split the integral up into two parts. We can write

$$\begin{aligned}
\mathbb{E}Y^D &\ll \frac{D!}{D^{D/2}e^{-D/2}D^{1/2}} \int_0^{\pi(D/2)^{1/2}} e^{-\psi^2/2}\left(1+O(\psi^4/D)\right)\mathrm{d}\psi \\
&\ll \frac{D!}{D^{D/2}e^{-D/2}D^{1/2}} \left(\int_0^{D^{1/4}} e^{-\psi^2/2}\mathrm{d}\psi + \int_{D^{1/4}}^{\pi(D/2)^{1/2}} De^{-(D^{1/2})/2}\mathrm{d}\psi\right),
\end{aligned}$$

and we get the desired result, since both integrals are bounded above by some absolute constant. □

Lemma 7.11. *We have for* $1 \leqslant d \leqslant (\log\log x)^{\sigma-1}$ *that*

$$\mathbb{E}\left(\frac{T_x - L'\mu_q}{\sqrt{L'\sigma_q^2}}\right)^d = \mathbb{E}\left(\frac{\overline{T}_x - L'\mu_q}{\sqrt{L'\sigma_q^2}}\right)^d + F(x,d),$$

with

$$F(x,d) = O_\sigma\left(\left(\frac{q^2}{2\sigma_q}L^{1/2}\right)^d e^{-\tilde{c}(\log\log x)^\sigma}\right).$$

Proof. Using the definition of T_x and $\overline{T}_x$, we get

$$\begin{aligned}
&\mathbb{E}\left(T_x - L'\mu_q\right)^d - \mathbb{E}\left(\overline{T}_x - L'\mu_q\right)^d \\
&\quad= \sum_{(\log\log x)^\sigma \leqslant j_1,\dots,j_d \leqslant L-(\log\log x)^\sigma} \mathbb{E}\Big((D_{j_1,x}-\mu_q)\cdots(D_{j_d,x}-\mu_q) \\
&\qquad\qquad - (Z_{j_1}-\mu_q)\cdots(Z_{j_d}-\mu_q)\Big),
\end{aligned}$$

where $D_{j,x}$ denotes the random variable defined by $D_{j,x}(m) = \varepsilon_j(m)$. A standard change of variables argument (see for example [Bil86, Theorem 16.12]) shows that this is the same as

$$\sum_{(\log\log x)^\sigma \leqslant j_1,\dots,j_d \leqslant L-(\log\log x)^\sigma} \quad \sum_{0\leqslant \ell_{j_1},\dots,\ell_{j_d}<q} (\ell_{j_1}-\mu_q)\cdots(\ell_{j_d}-\mu_q)$$
$$\cdot \Big(\mathbf{Pr}\,(D_{j_1}=\ell_{j_1},\dots,D_{j_d}=\ell_{j_d}) - \mathbf{Pr}\,(Z_{j_1}=\ell_{j_1},\dots,Z_{j_d}=\ell_{j_d})\Big),$$

where $\mathbf{Pr}$ denotes our probability measure on $I(x,k)$. It follows from Lemma 7.6 that $\mathbf{Pr}\,(D_{j_1,x}=\ell_{j_1},\dots,D_{j_d,x}=\ell_{j_d})$ is the same as

$$\mathbf{Pr}\,(Z_{j_1}=\ell_{j_1},\dots,Z_{j_d}=\ell_{j_d}) + O_\sigma\Big(e^{-\tilde{c}(\log\log x)^\sigma}\Big)$$

for $(\log\log x)^\sigma \leqslant j_1,\dots,j_d \leqslant L-(\log\log x)^\sigma$. This shows the desired result. □

Proof of Proposition 7.5. In what follows, we will show that $\overline{T}_x$ is a good approximation of the (truncated) sum-of-digits function. In order to prove (7.7), it suffices to verify that

$$|\varphi_2(t)-\varphi_3(t)| \ll \frac{|t|}{\log x} \tag{7.11}$$

uniformly for real t with $|t| \ll \log\log x$. Using Taylor's theorem we have for every even integer $D>0$,

$$\begin{aligned}\mathbb{E}\,e^{itX} - \mathbb{E}\,e^{itY} &= \sum_{d<D}\frac{(it)^d}{d!}\Big(\mathbb{E}\,X^d - \mathbb{E}\,Y^d\Big)\\ &\quad + O\left(\frac{|t|^D}{D!}\left|\mathbb{E}\,|X|^D - \mathbb{E}\,|Y|^D\right| + 2\frac{|t|^D}{D!}\mathbb{E}\,|Y|^D\right)\\ &\ll |t|\max_{d\leqslant D}(|\mathbb{E}\,X^d - \mathbb{E}\,Y^d|)e^{|t|} + \frac{|t|^D}{D!}\mathbb{E}\,Y^D.\end{aligned}$$

Let $\varepsilon>0$ and set $\sigma=3+\varepsilon$. If we choose $D=\lfloor(\log\log x)^{2+\varepsilon/2}\rfloor$ (and assume without loss of generality that D is even), then Lemma 7.9 implies that

$$\frac{|t|^D}{D!}\mathbb{E}\,Y^D \ll_\varepsilon |t|/\log x.$$

As to complete the proof of Proposition 7.5, it remains to compare the moments of X and Y. Lemma 7.11 yields

$$|\mathbb{E}X^d - \mathbb{E}Y^d| \ll_\varepsilon \left(\frac{q^2}{2\sigma_q}L^{1/2}\right)^d e^{-\tilde{c}(\log\log x)^{3+\varepsilon}} \ll_\varepsilon e^{-c(\log\log x)^2},$$

with an appropriate constant $c>0$. This implies (7.11) and Proposition 7.5 is finally shown. □

7.2.2 Proof of Theorem 7.1

Since the integrand in (7.4) is periodic of period 1, we can shift the integral to the interval $[-1/(2(q-1)), 1-1/(2(q-1))]$. This allows us to reduce the integral to the interval $[-1/(2(q-1)), 1/(2(q-1))]$:

$$\begin{aligned}\#\{n \leqslant x : s_q(g(n)) = k\} =& \int_{-\frac{1}{2(q-1)}}^{1-\frac{1}{2(q-1)}} S(\alpha)\,\mathrm{e}(-\alpha k)\mathrm{d}\alpha\\ =&\sum_{\ell=0}^{q-2}\int_{-\frac{1}{2(q-1)}}^{\frac{1}{2(q-1)}} S\left(\alpha+\frac{\ell}{q-1}\right)\mathrm{e}\left(-\left(\alpha+\frac{\ell}{q-1}\right)k\right)\mathrm{d}\alpha\\ =&(q-1)\int_{-\frac{1}{2(q-1)}}^{\frac{1}{2(q-1)}} S_k(\alpha)\,\mathrm{e}(-\alpha k)\mathrm{d}\alpha,\end{aligned}$$

where

$$S_k(\alpha) = \sum_{n\in I(x,k)} \mathrm{e}(\alpha s_q(g(n))).$$

Now we consider the last integral separately whether α is small or big, namely

$$\int_{-\frac{1}{2(q-1)}}^{\frac{1}{2(q-1)}} = \int_{|\alpha|\leqslant(\log\log x)(\log x)^{-1/2}} + \int_{(\log\log x)(\log x)^{-1/2}<|\alpha|\leqslant 1/(2(q-1))}. \qquad (7.12)$$

The second integral (where α is big), can be bounded above using Condition (C2). Indeed, it also implies the upper bound

$$S_k(\alpha) \ll x^{1-c_1\|(q-1)\alpha\|^2}(\log x)^{d_1},$$

and we obtain for the second integral in (7.12) the estimate

$$\int S_k(\alpha)\ \mathrm{e}(-\alpha k)\,d\alpha \ll x\,x^{-c_1(q-1)^2(\log\log x)^2(\log x)^{-1}}(\log x)^{d_1} \ll \frac{x}{\log x}.$$

This expression is bounded by the error term stated in the theorem. For the upper bound of the first integral in (7.12), we use Proposition 7.5. We have

$$\begin{aligned}&\int_{|\alpha|\leqslant\log\log x(\log x)^{-1/2}} S_k(\alpha)\,\mathrm{e}(-\alpha k)\,\mathrm{d}\alpha\\ &\quad= R(x,k)\int_{|\alpha|\leqslant\log\log x(\log x)^{-1/2}} \mathrm{e}(\alpha(\mu_q\log_q g(x)-k))\,e^{-2\pi^2\alpha^2\sigma_q^2\log_q g(x)}\mathrm{d}\alpha\\ &\qquad+ R(x,k)\int_{|\alpha|\leqslant\log\log x(\log x)^{-1/2}} O(|\alpha|(\log\log x)^{3+\varepsilon})\,\mathrm{d}\alpha.\end{aligned}$$

Next, we use the substitution $\alpha = t/(2\pi\sigma_q(\log_q g(x))^{1/2})$ and obtain

$$\frac{R(x,k)}{2\pi\sigma_q(\log_q g(x))^{1/2}} \int_{-\infty}^{\infty} e^{it\Delta_k - t^2/2}\,\mathrm{d}t + O\left(\frac{R(x,k)}{(\log g(x))^{1/2}} \int_{|t|\geqslant \log\log x} e^{-t^2/2}\mathrm{d}t\right)$$
$$+ O\left(R(x,k)\frac{(\log\log x)^{5+\varepsilon}}{\log x}\right)$$
$$= \frac{R(x,k)}{\sqrt{2\pi\sigma_q^2 \log_q g(x)}}\left(e^{-\Delta_k^2/2} + O\Big(e^{-(\log\log x)^2/2}\Big) + O\left(\frac{(\log\log x)^{5+\varepsilon}}{(\log x)^{\frac{1}{2}}}\right)\right),$$

where $\Delta_k = \frac{k-\mu_q \log_q g(x)}{\sigma_q(\log_q g(x))^{1/2}}$. Since the first O-term is bounded by the second one, Theorem 7.1 follows.

7.3 The sum of digits in $\mathbb{Z}[i]$

In this section we prove Theorem 7.3. Let $q = -a \pm i$ for some $a \geqslant 1$. In what follows, we write s_q for the complex sum-of-digits function $s_q^{\mathbb{C}}$ since we are only dealing with Gaussian integers. Let $k \geqslant 0$ and set $I(N,k) = \{z \in \mathcal{C}_N : g(z) \equiv k \bmod q-1\}$. As in the real case we have $\{z \in \mathcal{C}_N : s_q(g(z)) = k\} \subseteq I(N,k)$ and the desired result holds trivially if $g_k = 0$. Thus we can assume that $g_k > 0$. In this case, we have to show

$$\#\{z \in \mathcal{C}_N : s_q(g(z)) = k\} = \frac{(a^2+2a+2)R(N,k)}{\sqrt{2\pi\sigma_Q^2 \log_Q |g(\lfloor\sqrt{N}\rfloor)|^2}}\left(e^{-\Delta_k^2/2} + O\left(\frac{(\log\log N)^{11}}{(\log N)^{1/2}}\right)\right),$$

where $R(N,k) = \#I(N,k)$. The implied constant in the O-term may depend on the function g and on the base q, and we omit the dependence in the notation.

We have $s_q(z) \equiv \Re(z) \pm (a+1)\Im(z) \bmod (a^2+2a+2)$ (see Proposition A.4) and $s_q(z) \equiv z \bmod (q-1)$. With Lemma 2.22 we can characterize $I(N,k)$ in two different ways, namely,

$$\begin{aligned} I(N,k) &= \{z \in \mathcal{C}_N : s_q(g(z)) \equiv k \bmod \left(a^2+2a+2\right)\} \\ &= \{z \in \mathcal{C}_N : \Re(g(z)) \pm (a+1)\Im(g(z)) \equiv k \bmod (a^2+2a+2)\}. \end{aligned}$$

Proposition 7.12. *Let $q = -a \pm i$, where $a \geqslant 1$. Then we have for every non-negative integer k,*

$$\sum_{z\in I(N,k)} \mathrm{e}\left(\alpha s_q(g(z))\right) = R(N,k)\,\mathrm{e}\left(\alpha\mu_Q \log_Q |g(\lfloor\sqrt{N}\rfloor)|^2\right)$$
$$\cdot \left(e^{-2\pi^2\alpha^2\sigma_Q^2 \log_Q |g(\lfloor\sqrt{N}\rfloor)|^2} + O\left(|\alpha|(\log\log N)^9\right)\right) \quad (7.13)$$

uniformly for real α with $|\alpha| \leqslant (\log\log N)(\log N)^{-1/2}$.

We translate again this statement into a probabilistic language. Condition ($\mathbb{C}$1) implies that there exists a constant ξ such that the Gaussian integers $g(z)$ have less than or equal to

$$L := (\log_Q |g(\lfloor\sqrt{N}\rfloor)|^2) + \xi$$

digits for all $z \in \mathcal{C}_N$. In what follows, we will prove that

$$\sum_{z\in I(N,k)} \mathrm{e}\left(\alpha s_q\left(g(z)\right)\right) = R(N,k)\,\mathrm{e}\left(\alpha\mu_Q L\right) \cdot \left(e^{-2\pi^2\alpha^2\sigma_Q^2 L} + O\left(|\alpha|(\log L)^9\right)\right)$$

uniformly for real α with $|\alpha| \leqslant 2(\log L)L^{-1/2}$. This implies Proposition 7.12. If we assume that every number in the set $I(N,k)$ is equally likely, then the function which assigns each number its j-th digit is a random variable. Hence, the sum-of-digits function $S_N(z) := s_q(g(z))$ can also be interpreted as a random variable. Thus we have to show (compare with (7.7)) that the characteristic function $\varphi_1(t)$ of $(S_N - L\mu_Q)/(L\sigma_Q^2)^{1/2}$ satisfies

$$\varphi_1(t) = e^{-t^2/2} + O\left(|t|\frac{(\log L)^9}{L^{\frac{1}{2}}}\right) \tag{7.14}$$

uniform for $|t| \leqslant 4\pi\sigma_Q L^{1/2}$.

7.3.1 Joint distribution and fundamental domain

As in Section 7.2, we truncate the sum-of-digits function and approximate it appropriately. Let Z_j be a sequence of independent random variables with range $\{0, 1, \ldots, Q-1\}$ and uniform probability distribution, and set

$$T_N := \sum_{(\log L)^9\leqslant j\leqslant L-(\log L)^9} \varepsilon_j(g(z)), \qquad \overline{T}_N := \sum_{(\log L)^9\leqslant j\leqslant L-(\log L)^9} Z_j.$$

Furthermore, set

$$L' = \#\left\{j \in \mathbb{Z} : (\log L)^9 \leqslant j \leqslant L - (\log L)^9\right\} = L - 2(\log L)^9 + O(1).$$

Define the random variables X and Y by $X := (T_N - L'\mu_Q)/(L'\sigma_Q^2)^{1/2}$ and $Y := (\overline{T}_N - L'\mu_Q)/(L'\sigma_Q^2)^{1/2}$, and let $\varphi_2(t)$ be the characteristic function of X and $\varphi_3(t)$ the characteristic function of Y.

In this section we compare the joint distribution of the summands of T_N and $\overline{T}_N$. Although the result of the following lemma is quite similar to the corresponding result in the real case (compare with Lemma 7.6), the given proof is completely different since we have to deal with the special structure of the fundamental domain of the base-q representation system in $\mathbb{C}$.

Lemma 7.13. *Let $1 \leqslant d \leqslant L'$ and $j_1, j_2, \dots, j_d$ and $\ell_1, \ell_2, \dots, \ell_d$ integers with*

$$(\log L)^9 \leqslant j_1 < j_2 < \cdots < j_d \leqslant L - (\log L)^9$$

and $\ell_1, \ell_2, \dots, \ell_d \in \{0, 1, \dots, Q-1\}$. Then we have uniformly

$$\frac{1}{R(N,k)} \# \{z \in I(N,k) : \varepsilon_{j_1}(g(z)) = \ell_1, \dots, \varepsilon_{j_d}(g(z)) = \ell_d\}$$
$$= Q^{-d} + O\left(de^{-\gamma_2 r} + e^{\gamma_3 r d - \gamma_1 (\log L)^9}\right),$$

where γ_1, γ_2 and γ_3 are positive constants and $r > 0$ is an arbitrary integer.

We adopt the notion of Gittenberger and Thuswaldner [GT00, Chapter 3] (see also Appendix A.1) and define the fundamental domain of the base-q representation system by

$$\mathcal{F}' = \left\{ z \in \mathbb{C} : z = \sum_{j=1}^{\infty} \varepsilon_j(z) q^{-j}, \varepsilon_j \in \mathcal{N} \right\}.$$

Figure 7.1 shows the fundamental domain in base $-1+i$ and it is called the twindragon fractal (see [Knu81]). Different colors correspond to numbers with different leading digits. For example, the base q-representation of complex numbers lying in the black region start with $0.1\dots$, while in the region with the darkest gray coloring the numbers start with $0.01\dots$. Every complex number z can be represented as $z = \alpha_0 + \alpha_1 q$ with unique real numbers α_0 and α_1. Thus, the mapping

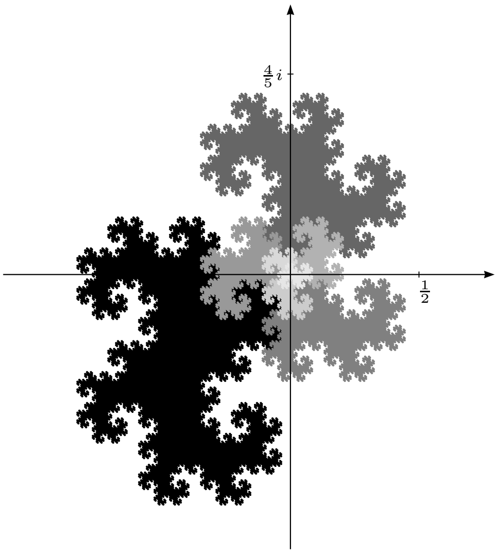


Figure 7.1: Twindragon

$$\varphi : \mathbb{C} \to \mathbb{R}^2, \quad z = \alpha_0 + \alpha_1 q \mapsto (\alpha_0, \alpha_1)$$

is well defined and it is called the φ-embedding of $\mathcal{F}'$ in $\mathbb{R}^2$. We set

$$\mathcal{F} := \varphi(\mathcal{F}') = \left\{ z \in \mathbb{R}^2 : z = \sum_{j=1}^{\infty} E^{-j} \varepsilon_j(z), \varepsilon_j \in \varphi(\mathcal{N}) \right\},$$

with

$$E = \begin{pmatrix} 0 & -1-a^2 \\ 1 & -2a \end{pmatrix}.$$

Note that $\varphi(qz) = E\varphi(z)$. If $m \in \mathcal{N}$, we will have to deal with the domain containing all the numbers whose fractional parts start with the digit m (for example, the gray regions in Figure 7.1 for $m = 0$ and the black region for $m = 1$). We denote the embedded version by

$$\mathcal{F}_m = E^{-1} \left(\mathcal{F} + \varphi(m) \right).$$

Since this region has a rather complicated shape, we have to approximate it. Therefore we use the following lemma which is proved in [GT00, Lemma 3.1].

Lemma 7.14. *For all $m \in \mathcal{N}$ and all $r \in \mathbb{N}$ there exists an axially parallel tube $P_{r,m}$ (a union of axially parallel rectangles) with the following properties:*

(i) $\partial \mathcal{F}_m \subset P_{r,m}$ *for all* $r \in \mathbb{N}$.

(ii) $\lambda_2(P_{r,m}) = O(\mu^r / Q^r)$.

(iii) $P_{r,m}$ *consists of* $O(\mu^k)$ *axially parallel rectangles, each of which has Lebesgue measure* $O(Q^{-r})$.

The constant μ satisfies $1 < \mu < Q$.

In the proof of the lemma Gittenberger and Thuswaldner constructed a polygon $\Pi_{r,m}$ with axes-parallel sides such that

$$P_{r,m} = \{ z \in \mathbb{R}^2 : \| z - \Pi_{r,m} \|_\infty \leqslant \tilde{c}\, |q|^{-r} \},$$

where $\tilde{c}$ is an absolute constant that can be chosen $\geqslant 1$. For the remaining part of this section we fix to each pair (r, m) the polygon $\Pi_{r,m}$, the corresponding tube $P_{r,m}$ and denote by $I_{r,m}$ the set of all points inside $\Pi_{r,m}$. We define

$$f_m(x, y) = \frac{1}{\Delta^2} \int_{-\Delta/2}^{\Delta/2} \int_{-\Delta/2}^{\Delta/2} \Psi_m(x + x_1, y + y_1) \mathrm{d}x_1 \mathrm{d}y_1,$$

where $\Delta = \tilde{c}\, |q|^{-r}$ and

$$\Psi_m(x, y) = \begin{cases} 1, & \text{if } (x, y) \in I_{r,m}, \\ 1/2, & \text{if } (x, y) \in \Pi_{r,m}, \\ 0, & \text{otherwise.} \end{cases}$$

The function f_m is a so-called Urysohn function which equals 1 for $(x, y) \in I_{r,m} \backslash P_{r,m}$, 0 for $(x, y) \in \mathbb{R}^2 \setminus (I_{r,m} \cup P_{r,m})$ and is an interpolation of these values in between. The next lemma gives estimates for the Fourier coefficients of this function and can be found in [GT00, Lemma 3.2 and Lemma 3.3].

Lemma 7.15. *Let $f_m(x,y) = \sum_{n_1,n_2\in\mathbb{Z}} c_{n_1,n_2}\,\mathrm{e}(n_1x+n_2y)$ be the Fourier expansion of f_m. Then for the Fourier coefficients c_{n_1,n_2} we get the estimates*

$$\begin{aligned} c_{n_1,n_2} &= O\left(\frac{\mu^r}{\Delta^2 n_1^2 n_2^2}\right) \qquad (n_1,n_2\neq 0),\\ c_{n_1,0} &= O\left(\frac{\mu^r}{\Delta n_1^2}\right) \qquad (n_1\neq 0),\\ c_{0,n_2} &= O\left(\frac{\mu^r}{\Delta n_2^2}\right) \qquad (n_2\neq 0),\\ c_{0,0} &= \frac{1}{Q}. \end{aligned}$$

Furthermore we have for $n_1,n_2\neq 0$ that $c_{n_1,n_2}=0$ if $q \mid (\bar{q}n_1-n_2)$.

Before we start proving Lemma 7.13, we need an auxiliary result. We set

$$F_j = \#\left\{z\in I(N,k) : \varphi\left(\frac{g(z)}{q^{j+1}}\right) \in \bigcup_{m\in\mathcal{N}} P_{r,m} \bmod \mathbb{Z}^2\right\}.$$

Lemma 7.16. *We have uniformly for* $(\log L)^9 \leqslant j \leqslant L-(\log L)^9$,

$$\frac{1}{R(N,k)}F_j \ll \left(\frac{\mu}{Q}\right)^r + \mu^r e^{-\gamma_1(\log L)^9},$$

where γ_1 is a positive constant.

Proof. From Lemma 7.14 it follows that we can subdivide each tube $P_{r,m}$ into a family of $O(\mu^r)$ rectangles (which have Lebesgue measure $O(Q^{-r})$) such that we have

$$\frac{1}{R(N,k)}F_j \leqslant \frac{1}{R(N,k)}\sum_{m\in\mathcal{N}}\sum_{G_m\subseteq P_{r,m}} F_j(G_m), \tag{7.15}$$

where $F_j(G_m)$ is defined by

$$F_j(G_m) = \#\left\{z\in I(N,k) : \varphi\left(\frac{g(z)}{q^{j+1}}\right) \in G_m \bmod \mathbb{Z}^2\right\},$$

and the second sum in (7.15) runs over all rectangles G_m in which we subdivide $P_{r,m}$. In what follows we show that there exists a positive constant γ_1 such that the discrepancy D of the sequence $\varphi\left(\frac{g(z)}{q^{j+1}}\right)$ where $z\in I(N,k)$ is bounded by $D \ll e^{-\gamma_1(\log L)^9}$. This then implies

$$\frac{1}{R(N,k)}F_j(G_m) \ll \lambda_2(G_m) + D \ll \frac{1}{Q^r} + e^{-\gamma_1(\log L)^9},$$

and the result of the lemma follows (see Lemma 7.14). This leads us to the 2-dimensional Erdős-Turán-Koksma inequality (see Theorem A.19). We have

$$D \ll \frac{2}{H+1} + \sum_{0<\|h\|_\infty\leqslant H}\frac{1}{r(h)}\left|\frac{1}{R(N,k)}\sum_{z\in I(N,k)} \mathrm{e}\left(h\cdot\varphi\left(\frac{g(z)}{q^{j+1}}\right)\right)\right|, \tag{7.16}$$

where $r(h)=\prod_{1\leqslant i\leqslant k}\max(1,|h_i|)$ for $h=(h_1,\dots,h_k)\in\mathbb{Z}^k$. It is easy to see that

$$\tau(z):=\left(\frac{1}{2}\operatorname{tr}(z),\frac{1}{2}\operatorname{tr}(qz)\right)^T=X\varphi(z)\quad\text{with}\quad X:=\begin{pmatrix}1 & -a\\ -a & a^2-1\end{pmatrix}.$$

Then we have

$$h\cdot\varphi\left(\frac{g(z)}{q^{j+1}}\right)=h^TX^{-1}\tau\left(\frac{g(z)}{q^{j+1}}\right)=\frac{1}{2}\operatorname{tr}\left(\left(\frac{\tilde{h}_1}{q^{j+1}}+\frac{\tilde{h}_2}{q^j}\right)g(z)\right),$$

where $(\tilde{h}_1,\tilde{h}_2):=h^TX^{-1}$. Note furthermore, that $2(\Re(g(z))+(a+1)\Im(g(z)))=\operatorname{tr}((1-i(a+1))g(z))$. Using the definition of $I(N,k)$ one easily obtains for the inner sum in (7.16),

$$\sum_{z\in I(N,k)}\mathrm{e}\left(h\cdot\varphi\left(\frac{g(z)}{q^{j+1}}\right)\right)=\frac{1}{a^2+2a+2}\sum_{\ell=0}^{a^2+2a+1}\mathrm{e}\left(-\frac{k\ell}{a^2+2a+2}\right)$$
$$\cdot\sum_{z\in\mathcal{C}_N}\mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\left(\frac{\tilde{h}_1}{q^{j+1}}+\frac{\tilde{h}_2}{q^j}+\frac{\ell(1-i(a+1))}{a^2+2a+2}\right)g(z)\right)\right).$$

We set

$$\frac{A}{B}=\frac{\tilde{h}_1}{q^{j+1}}+\frac{\tilde{h}_2}{q^j}+\frac{\ell(1-i(a+1))}{a^2+2a+2},$$

with $(A,B)=1$. Then $|q|^{j-2}/H\leqslant|B|\leqslant|q|^{j+4}$ (note, that $\tilde{h}_1,\tilde{h}_2\leqslant Q^{3/2}H$). If we set $H=\lfloor Q^{\frac{1}{3}(\log L)^9}\rfloor$, we have

$$Q^{\frac{2}{3}(\log L)^9}\ll|B|^2\ll|g(\lfloor\sqrt{N}\rfloor)|^2Q^{-\frac{2}{3}(\log L)^9}.$$

Condition ($\mathbb{C}$3) implies

$$\begin{aligned}D&\ll\frac{1}{H}+\frac{N}{R(N,k)}N^{-c_2(\log\log N)^9/\log N}(\log N)^{d_2}\sum_{0\leqslant\|h\|_\infty\leqslant H}\frac{1}{r(h)}\\&\ll\frac{1}{H}+e^{-c_2(\log\log N)^9}(\log N)^{d_2}(\log H)^2\\&\ll e^{-\gamma_1(\log L)^9},\end{aligned}$$

where γ_1 is a suitable positive constant. Here we used that $N/R(N,k)\ll 1$, which follows from Condition ($\mathbb{C}$1) and the properties of the sets $\mathcal{C}_N$, $N\geqslant 0$. This concludes the proof of Lemma 7.16. □

Proof of Lemma 7.13. We set

$$t_{\mathbf{l},\mathbf{j}}(z)=\prod_{h=1}^{d}f_{\ell_h}\left(\varphi\left(\frac{z}{q^{j_h+1}}\right)\right),$$

where $\mathbf{l} = (\ell_1, \ldots, \ell_d)$ and $\mathbf{j} = (j_1, \ldots, j_d)$. The following fundamental relation allows us to use our just obtained results. We have

$$\begin{aligned}\left| \# \left\{ z \in I(N,k) : \varepsilon_{j_1}\left(g(z)\right) = \ell_1, \ldots, \varepsilon_{j_d}\left(g(z)\right) = \ell_d \right\} - \sum_{z \in I(N,k)} t_{\mathbf{l},\mathbf{j}}\left(g(z)\right) \right| \\ \leqslant F_{j_1} + \cdots + F_{j_d} \\ \leqslant dR(N,k) \left(\left(\frac{\mu}{Q} \right)^r + \mu^r e^{-\gamma_1 (\log L)^9} \right).\end{aligned}$$

Thus it remains to study the sum $\sum_{z \in I(N,k)} t_{\mathbf{l},\mathbf{j}}\left(g(z)\right)$. It is easy to see that

$$t_{\mathbf{l},\mathbf{j}}(g(z)) = \sum_{M \in \mathcal{M}} T_M \, \mathrm{e} \left(\sum_{h=1}^{d} \mu_h \varphi \left(\frac{g(z)}{q^{j_h+1}} \right) \right),$$

where $\mathcal{M} = \{M = (\mu_1, \ldots, \mu_d) : \mu_h = (m_{h1}, m_{h2})$ with $m_{h1}, m_{h2} \in \mathbb{Z}; h = 1, \ldots, d\}$, and $T_M = \prod_{h=1}^{d} c_{m_{h1}, m_{h2}}$. Hence we can write

$$\sum_{z \in I(N,k)} t_{\mathbf{l},\mathbf{j}}(g(z)) = \sum_{M \in \mathcal{M}} T_M \sum_{z \in I(N,k)} \mathrm{e} \left(\sum_{h=1}^{d} \mu_h \varphi \left(\frac{g(z)}{q^{j_h+1}} \right) \right).$$

If $M = 0$, then $T_M = Q^{-d}$ (by Lemma 7.15). If $M = (\mu_1, \ldots, \mu_d) \neq 0$ such that there exists an integer h with $q \mid \bar{q} m_{h1} - m_{h2}$ then $T_M = 0$ (again by Lemma 7.15). In all other cases we have (using the same notation as in the proof of Lemma 7.16)

$$\begin{aligned}\sum_{h=1}^{d} \mu_h \varphi \left(\frac{g(z)}{q^{j_h+1}} \right) &= \sum_{h=1}^{d} \mu_h X^{-1} \tau \left(\frac{g(z)}{q^{j_h+1}} \right) \\ &= \frac{1}{2} \operatorname{tr} \left(\mp i \sum_{h=1}^{d} \frac{\bar{q} m_{h1} - m_{h2}}{q^{j_h+1}} g(z) \right),\end{aligned}$$

where the choice of the sign depends on the sign of $q = -a \pm i$. If we set

$$\frac{A'}{B'} = \mp i \sum_{h=1}^{d} \frac{\bar{q} m_{h1} - m_{h2}}{q^{j_h+1}},$$

where $(A', B') = 1$, we see that there exists a constant c' such that $|q|^{c' j_1} \leqslant |B'| \leqslant |q|^{j_d+1}$ (compare with [DMR09, Proof of Lemma 4.5] and note, that $q \nmid \bar{q} m_{h1} - m_{h2}$). Furthermore we have

$$\begin{aligned}\sum_{z \in I(N,k)} t_{\mathbf{l},\mathbf{j}}(g(z)) = \frac{1}{a^2+2a+2} \sum_{r=0}^{a^2+2a+1} \mathrm{e} \left(-\frac{rk}{a^2+2a+2} \right) \\ \cdot \sum_{M \in \mathcal{M}} T_M \sum_{z \in \mathcal{C}_N} \mathrm{e} \left(\frac{1}{2} \operatorname{tr} \left(\left(\frac{A'}{B'} + \frac{r(1-i(a+1))}{a^2+2a+2} \right) g(z) \right) \right).\end{aligned}$$

If we write the inner sum as $\sum_{z\in\mathcal{C}_N} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{A}{B}g(z)\right)\right)$, we see that $|B'| \leqslant |B| \ll |B'|$ since $(q, a^2+2a+2) = 1$. Thus we have $Q^{c'(\log L)^9} \ll |B|^2 \ll |g(\lfloor\sqrt{N}\rfloor)|^2 Q^{-(\log L)^9}$ and Condition ($\mathbb{C}3$) yields

$$\sum_{z\in I(N,k)} t_{\mathbf{l},\mathbf{j}}(g(z)) \ll \frac{R(N,k)}{Q^d} + O\left(N^{1-c_2(\log\log N)^9/\log N}(\log N)^{d_2}\sum_{M\neq 0}|T_M|\right),$$

where we assume without loss of generality that the constants c_2 and d_2 are the same as in the proof of the last lemma. It is easy to see that $\sum_M |T_M| \ll \left(\frac{\mu^r}{\Delta^2}\right)^d \ll (\mu^r Q^r)^d$. Thus we have

$$\sum_{z\in I(N,k)} t_{\mathbf{l},\mathbf{j}}(g(z)) \ll \frac{R(N,k)}{Q^d} + O\left(R(N,k)e^{-c_2(\log\log N)^9+\log(\mu Q)rd}(\log N)^{d_2}\right).$$

Finally, we set $\gamma_2 = -\log(\mu/Q)$, $\gamma_3 = \log(\mu Q)$ and we obtain the desired result. □

7.3.2 Proofs of Proposition 7.12 and Theorem 7.3

The proof of Proposition 7.12 as well as the final steps in the proof of Theorem 7.3 are similar to the real case. Hence, we give here only a rough outline.

Proof of Proposition 7.12. Recall that

$$T_N = \sum_{(\log L)^9 \leqslant j \leqslant L-(\log L)^9} \varepsilon_j(g(n)), \qquad \overline{T}_N = \sum_{(\log L)^9 \leqslant j \leqslant L-(\log L)^9} Z_j,$$

$L' = \#\left\{j\in\mathbb{Z} : (\log L)^9 \leqslant j \leqslant L-(\log L)^9\right\}$, and X and Y are defined by $X = (T_N - L'\mu_Q)/(L'\sigma_Q^2)^{1/2}$ and $Y := (\overline{T}_N - L'\mu_Q)/(L'\sigma_Q^2)^{1/2}$. Note furthermore, that $\varphi_2(t)$ is the characteristic function of X and $\varphi_3(t)$ the characteristic function of Y. It is easy to see that the statement of Lemma 7.7 holds also true in the Gaussian integers. We have

$$|\varphi_1(t) - \varphi_2(t)| \ll |t|\frac{(\log L)^9}{L^{1/2}}.$$

Furthermore, $\varphi_3(t)$ can be approximated by

$$\varphi_3(t) = e^{-t^2/2}\left(1 + O\left(\frac{t^4}{L}\right)\right)$$

whenever $|t| \leqslant L^{\frac{1}{4}}$ (this is Lemma 7.8 with q replaced by Q). Hence, in order to prove (7.14) it suffices to show that we have uniformly for real t with $|t| \ll \log L$,

$$|\varphi_2(t) - \varphi_3(t)| = O\left(\frac{|t|}{L}\right). \tag{7.17}$$

As in the real case, Taylor's theorem implies for every even integer $D > 0$,

$$\mathbb{E}\, e^{itX} - \mathbb{E}\, e^{itY} \ll |t| \max_{d \leqslant D}(|\mathbb{E}\, X^d - \mathbb{E}\, Y^d|)e^{|t|} + \frac{|t|^D}{D!}\mathbb{E}\, Y^D.$$

We have (see Lemma 7.9)

$$\mathbb{E}\, Y^D \ll \frac{D!}{D^{D/2}e^{-D/2}D^{1/2}},$$

whenever $D = o((\log x)^{1/2})$. If we choose $D = \lfloor (\log L)^3 \rfloor$ (and assume without loss of generality that D is even), then we get

$$\frac{|t|^D}{D!}\mathbb{E}\, Y^D \ll |t|/L.$$

In order to complete the proof of Proposition 7.12, it remains to compare the moments of X and Y. Lemma 7.13 implies (the reasoning is the same as in the proof of Lemma 7.11)

$$|\mathbb{E}X^d - \mathbb{E}Y^d| \ll \left(\frac{Q^2}{2\sigma_Q}L^{1/2}\right)^d \left(de^{-\gamma_2 r} + e^{\gamma_3 rd - \gamma_1(\log L)^9}\right).$$

We choose $r = \lfloor (\log L)^5 \rfloor$ and obtain

$$\max_{d \leqslant D} |\mathbb{E}\, X^d - \mathbb{E}\, Y^d| \ll_{c,q} e^{-(\log L)^2}.$$

This shows (7.17) and the desired result follows. □

Proof of Theorem 7.3. With the help of the already proved results, the last steps of the proof of Theorem 7.3 work exactly the same way as the steps shown in Section 7.2.2. First, we use the periodicity of the integrand to obtain

$$\int_0^1 S(\alpha)\, \mathrm{e}(-\alpha k) d\alpha = (a^2 + 2a + 2) \int_{-\frac{1}{2(a^2+2a+2)}}^{\frac{1}{2(a^2+2a+2)}} S_k(\alpha)\, \mathrm{e}(-\alpha k) \mathrm{d}\alpha,$$

where

$$S_k(\alpha) = \sum_{z \in I(N,k)} \mathrm{e}(\alpha s_q(g(z))).$$

We consider the last integral separately whether α is small or big, namely

$$\int_{-\frac{1}{2(a^2+2a+2)}}^{\frac{1}{2(a^2+2a+2)}} = \int_{|\alpha| \leqslant (\log\log N)(\log N)^{-1/2}} + \int_{(\log\log N)(\log N)^{-1/2} < |\alpha| \leqslant 1/(2(a^2+2a+2))}. \tag{7.18}$$

Condition ($\mathbb{C}$2) implies

$$S_k(\alpha) \ll N^{1-c_1\|(a^2+2a+2)\alpha\|^2}(\log N)^{d_1},$$

and we obtain for the second integral in (7.18) the estimate

$$\int S_k(\alpha)\ \mathrm{e}(-\alpha k)\,d\alpha \ll N^{1-c_1\|(a^2+2a+2)\alpha\|^2}(\log N)^{d_1} \ll \frac{N}{\log N}.$$

Since $\#\mathcal{C}_N \approx N$, this expression is bounded by the error term stated in the theorem. For the first integral in (7.18), we use Proposition 7.12 and the same calculations as in Section 7.2.2 show that

$$\int_{|\alpha|\leqslant \log\log N(\log N)^{-1/2}} S_k(\alpha)\,\mathrm{e}(-\alpha k)\,\mathrm{d}\alpha$$
$$= \frac{R(N,k)}{\sqrt{2\pi\sigma_Q^2 \log_Q |g(\lfloor\sqrt{N}\rfloor)|^2}}\left(e^{-\Delta_k^2/2} + O\Big(e^{-(\log\log N)^2/2}\Big) + O\bigg(\frac{(\log\log N)^{11}}{(\log N)^{\frac{1}{2}}}\bigg)\right),$$

where $\Delta_k = \frac{k-\mu_Q \log_Q |g(\lfloor\sqrt{N}\rfloor)|^2}{\sqrt{\sigma_Q^2(\log_Q |g(\lfloor\sqrt{N}\rfloor)|^2}}$. This proves Theorem 7.3. □

7.4 Examples

We start with an example that can be easily deduced from the results presented in this chapter. Recall that Fouvry and Mauduit proved that

$$\#\{n < x : s_q(n) = \mu_q\lfloor\log_q n\rfloor + b(\lfloor\log_q n\rfloor)\} = \sqrt{\frac{6}{\pi(q^2-1)}}\frac{x}{\sqrt{\log_q x}} + O_K\left(\frac{x}{\log x}\right)$$

uniformly for any $x \geqslant 2$ and any $b : \mathbb{N} \to \mathbb{R}$ such that $|b(\nu)| \leqslant Kv^{1/4}$ and $\mu_q\nu + b(\nu) \in \mathbb{N}$ for any $n \geqslant 1$. The following slightly weaker result is an immediate consequence of Theorem 7.1:

Lemma 7.17. *We have, as $x \to \infty$,*

$$\#\{n < x : s_q(n) = \lfloor\mu_q \log_q n\rfloor\} = \sqrt{\frac{6}{\pi(q^2-1)}}\frac{x}{\sqrt{\log_q x}} + O_\varepsilon\left(x\frac{(\log\log x)^{5+\varepsilon}}{\log x}\right),$$

where $\varepsilon > 0$ is arbitrary but fixed.

Proof. The function $g(n) = n$ clearly satisfies Condition (C1) and the real numbers g_k are equal to $1/(q-1)$. Condition (C2) is a consequence of Gelfond's treatment of the sum-of-digits function (see [Gel68]) and the sums considered in Condition (C3) are geometric sums and it is easy to show that this condition holds also true. Thus we get

$$\#\{n < x : s_q(n) = k\} = \frac{x}{\sqrt{2\pi\sigma_q^2 \log_q x}}\left(e^{-\frac{\Delta_k^2}{2}} + O_\varepsilon\left(\frac{(\log\log x)^{5+\varepsilon}}{(\log x)^{1/2}}\right)\right), \qquad (7.19)$$

where $\Delta_k = \frac{k-\mu_q \log_q x}{\sqrt{\sigma_q^2 \log_q x}}$. Set $A_m(x) = \#\{n < x : s_q(n) = m\}$ and note that $\lfloor \mu_q \log_q n \rfloor = m$ if and only if $q^{m/\mu_q} \leqslant n < q^{(m+1)/\mu_q}$. Hence

$$\begin{aligned}\#\{n < x : s_q(n) = \lfloor \mu_q \log_q n \rfloor\} = &\sum_{m<\lfloor \mu_q \log_q x\rfloor} \Big(A_m(q^{(m+1)/\mu_q}) - A_m(q^{m/\mu_q}) \Big) \\ &+ A_{\lfloor \mu_q \log_q x\rfloor}(x) - A_{\lfloor \mu_q \log_q x\rfloor}(q^{\lfloor \mu_q \log_q x\rfloor/\mu_q}).\end{aligned} \tag{7.20}$$

Equation (7.19) implies that for $j \in \{0, 1\}$ we have

$$A_m(q^{(m+j)/\mu_q}) = \frac{1}{\sqrt{2\pi\sigma_q^2}} \frac{q^{(m+j)/\mu_q}}{(m/\mu_q)^{\frac{1}{2}}} \left(1 + O_\varepsilon \left(\frac{(\log m)^{5+\varepsilon}}{m^{1/2}} \right) \right).$$

For the first sum in (7.20), partial summation leads to (see for example [FM05, Lemma 2.3])

$$\begin{aligned}\frac{1}{\sqrt{2\pi\sigma_q^2}} &\sum_{m<\lfloor \mu_q \log_q x\rfloor} \left(\frac{q^{\frac{m+1}{\mu_q}}}{\sqrt{\frac{m}{\mu_q}}} - \frac{q^{\frac{m}{\mu_q}}}{\sqrt{\frac{m}{\mu_q}}} \right) \left(1 + O_\varepsilon \left(\frac{(\log m)^{5+\varepsilon}}{m^{1/2}} \right) \right) \\ &= \frac{q^{\frac{1}{\mu_q}} - 1}{\sqrt{2\pi\sigma_q^2}} \sum_{m<\lfloor \mu_q \log_q x\rfloor} \frac{q^{\frac{m}{\mu_q}}}{\sqrt{\frac{m}{\mu_q}}} \left(1 + O_\varepsilon \left(\frac{(\log m)^{5+\varepsilon}}{m^{1/2}} \right) \right) \\ &= \frac{1}{\sqrt{2\pi\sigma_q^2}} \frac{x}{\sqrt{\log_q x}} \left(q^{-\frac{\{\mu_q \log_q x\}}{\mu_q}} + O_\varepsilon \left(\frac{(\log\log x)^{5+\varepsilon}}{(\log x)^{1/2}} \right) \right).\end{aligned}$$

The second summand in (7.20) is

$$\begin{aligned}A_{\lfloor \mu_q \log_q x\rfloor}&(x) - A_{\lfloor \mu_q \log_q x\rfloor}(q^{\lfloor \mu_q \log_q x\rfloor/\mu_q}) \\ &= \frac{1}{\sqrt{2\pi\sigma_q^2}} \left(\frac{x}{\sqrt{\log_q x}} - \frac{q^{\frac{\lfloor \mu_q \log_q x\rfloor}{\mu_q}}}{\sqrt{\log_q x}} \right) \left(1 + O_\varepsilon \left(\frac{(\log\log x)^{5+\varepsilon}}{(\log x)^{1/2}} \right) \right)\end{aligned}$$

Summing these two equations up, we see that the periodic factor cancels out and Lemma 7.17 is shown. □

Remark 7.18. In the case of primes (that is, the function $g(n)$ assigns each n the n-th prime number), Condition (C1) is an immediate consequence of the prime number theorem and Condition (C3) is a well-known exponential sum estimate (see for example [IK04, Theorem 13.6]). Drmota, Mauduit, and Rivat [DMR09] proved that Condition (C2) is also valid. This shows that Theorem DMR (see Section 7.1) holds true even if we replace the error term $O((\log x)^{-1/2+\varepsilon})$ by $O((\log\log x)^{5+\varepsilon}/\log(x))$. The function $g(n) = n^2$ also satisfies Conditions (C1)–(C3). Condition (C1) holds trivially. A thorough analysis of the proof of [MR09, Theorem 1] shows that Condition (C2) holds true and Condition (C3) is again a well-known exponential sum estimate (see for example Appendix A.4 or [IK04, Theorem 8.1]). Furthermore, it is not hard to obtain a result similar to Lemma 7.17 for the sum of digits of primes (this was done by Drmota et al.) and of squares. However, in this case a periodic factor comes into play (compare with Corollary 7.23).

In the next section we show that the function $g(n) = \lfloor n^c \rfloor$ also satisfies these conditions if c is not an integer and if the base q is big enough. In the subsequent section we treat the sum of digits of squares in the Gaussian integers.

7.4.1 A local result for the sum of digits of $\lfloor n^c \rfloor$

Using the original method of Bassily and Kátai [BK95], it is relatively easy to show that $s_q(\lfloor n^c \rfloor)$ satisfies a central limit theorem. More precisely, we have

$$\frac{1}{x}\#\left\{n \leqslant x : s_q(\lfloor n^c \rfloor) \leqslant c\mu_q \log_q x + y\sqrt{\sigma_q^2 c \log_q x}\right\} = \Phi(y) + o(1), \tag{7.21}$$

where

$$\mu_q := \frac{q-1}{2}, \qquad \sigma_q^2 := \frac{q^2-1}{12},$$

and $\Phi(y)$ denotes the normal distribution function (see [DG]). In what follows we show a local version of this result if the base q is big enough.

Theorem 7.19. *Let $c > 0$ be a real number different from an integer. There exists a constant $q_0(c) \geqslant 2$ such that for all $q \geqslant q_0(c)$ the following holds: We have uniformly for all integers $k \geqslant 0$ and for any fixed $\varepsilon > 0$,*

$$\frac{1}{x}\#\{n \leqslant x : s_q(\lfloor n^c \rfloor) = k\} = \frac{1}{\sqrt{2\pi\sigma_q^2 c \log_q x}}\left(e^{-\Delta_k^2/2} + O_{c,q,\varepsilon}\left(\frac{(\log\log x)^{5+\varepsilon}}{(\log x)^{1/2}}\right)\right),$$

where $\Delta_k = \frac{k-\mu_q c \log_q x}{\sqrt{\sigma_q^2 c \log_q x}}$.

Remark 7.20. As in the linear case one can deduce from this theorem a result analog to Lemma 7.17.

Proof. We have to show that Conditions (C1)–(C3) hold true for the function $g(n) = \lfloor n^c \rfloor$. Condition (C1) is a consequence of Lemma 6.16. The validity of Condition (C2) is the content of Theorem 6.2. In order to prove Theorem 7.19 it remains to verify Condition (C3). Let $c > 0$ be a real number different from an integer. Furthermore, let $A, B \in \mathbb{Z}^+$ with $(A, B) = 1$ and $\sigma > 1$ such that $1 < B \leqslant x^c e^{-(\log\log x^c)^\sigma}$. Then we will show that

$$\sum_{1 \leqslant n \leqslant x} \mathrm{e}\left(\frac{A}{B}\lfloor n^c \rfloor\right) \ll_{c,\sigma} (\log x) x e^{-c'(\log\log x^c)^\sigma},$$

where $c' = \min\left(1, \frac{1}{c}\right)$. Let S be the considered sum. We start with the following estimate:

$$\left\|\frac{A}{B}\right\|^{-1} \leqslant B \leqslant x^c e^{-(\log\log x^c)^\sigma}. \tag{7.22}$$

If $0 < c < 1$, then we obtain (using the same calculations as in Section 6.4.1) that

$$\begin{aligned} S &\ll_c x^{1-c} \max_{1 \leqslant N \leqslant x^c} \left| \sum_{1 \leqslant m \leqslant N} \mathrm{e}\left(\frac{A}{B} m\right) \right| + x^{1-c} + x^c \\ &\ll_c x^{1-c} \frac{1}{\left|\sin \pi \frac{A}{B}\right|} + x^{1-c} + x^c. \end{aligned}$$

Using (7.22), this yields the desired result. Next, we treat the case $c > 1$. Let ν be the integer defined by $2^{\nu-1} < x \leqslant 2^\nu$. If x is sufficiently large, then

$$\nu_0 := \nu - \left\lfloor \frac{1}{c \log 2} (\log \log x^c)^\sigma \right\rfloor$$

is positive. Remark 6.14 implies

$$\begin{aligned} S &\leqslant 2^{\nu_0 - 1} + \sum_{\kappa=\nu_0}^{\nu} \sum_{\substack{2^{\kappa-1} < n \leqslant 2^\kappa \\ n \leqslant x}} \mathrm{e}\left(\frac{A}{B} \lfloor n^c \rfloor\right) \\ &\ll_{c,\sigma} 2^{\nu_0-1} + \sum_{\kappa=\nu_0}^{\nu} \left(\kappa 2^{\kappa(1-\tilde{\rho}/2)} + \frac{1}{\left\|\frac{A}{B}\right\|} 2^{\kappa(1-c)} \right). \end{aligned}$$

We finally obtain

$$\begin{aligned} S &\ll_{c,\sigma} 2^{\nu_0-1} + \nu q^{\nu(1-\tilde{\rho}/2)} + \nu x^c e^{-(\log\log x^c)^\sigma} 2^{\nu_0(1-c)} \\ &\ll_{c,\sigma} x e^{-\frac{1}{c}(\log\log x^c)^\sigma} + (\log x) x e^{-\frac{1}{c}(\log\log x^c)^\sigma}. \end{aligned}$$

Thus, Condition (C3) holds also true and Theorem 7.1 implies the desired result. □

7.4.2 A local result for the sum of digits of Gaussian squares

As already mentioned in Chapter 3, Gittenberger and Thuswaldner [GT00] dealt with the asymptotic normality of the sum-of-digits function of squares in the ring of Gaussian integers. (Actually, they considered polynomial subsequences.) They showed that

$$\frac{1}{\#\{z : |z|^2 < N\}} \# \left\{ |z|^2 < N : \frac{s_q^{\mathbb{C}}(z^2) - \mu_Q \log_Q N^2}{\sqrt{\sigma_Q^2 \log_Q N^2}} < y \right\} \to \Phi(y) \qquad (7.23)$$

as N goes to infinity and where Φ is the normal distribution function.

The next theorem provides asymptotic expansions for $\#\{z \in \mathcal{D}_N : s_q^{\mathbb{C}}(z^2) = k\}$ whenever $\mathcal{D}_N$ is a κ-$\mathbb{Z}[i]$ sequence with $\kappa = 1/2$ (see Chapter 3 for the definition of a κ-$\mathbb{Z}[i]$ sequence). If $\mathcal{D}_N$ is a disc with radius $\sqrt{N}$, it can be seen as a local version of (7.23). Recall that

$$\mathcal{A} = \{a \in \mathbb{N} : \text{ if } p \mid q = -a \pm i, \text{ then } |p| \geqslant \sqrt{689}\},$$

and that $Q(b, s) = \#\{z \in R_s : z^2 \equiv b \bmod s\}$, where R_s denotes a complete residue system modulo s.

Theorem 7.21. *Let $a \in \mathcal{A}$ and $(\mathcal{D}_N)_{N\in\mathbb{N}}$ be a 1/2-$\mathbb{Z}[i]$ sequence. Then we have uniformly for all integers $k \geqslant 0$ with $Q(k, q-1) \neq 0$,*

$$\frac{1}{\#\mathcal{D}_N}\#\left\{z \in \mathcal{D}_N : s_q^{\mathbb{C}}(z^2) = k\right\} = \frac{Q(k,q-1)}{\sqrt{2\pi\sigma_Q^2 \log_Q N^2}} \cdot e^{-\Delta_k^2/2} + O\left(\frac{(\log\log N)^{11}}{\log N}\right),$$

where $\Delta_k = \frac{k-\mu_Q \log_Q N^2}{\sqrt{\sigma_Q^2 \log_Q N^2}}$.

Remark 7.22. It would be interesting to obtain a corresponding local result for the sum of digits of primes in the Gaussian integers. Unfortunately, the exponential sum estimate shown in Chapter 2 (see Theorem 2.1) does not imply Condition ($\mathbb{C}$2) since this estimate is not uniform in α. This circumstance results from the fact that the implied constant of the symbol $\ll$ in the carry lemma (Lemma 2.17) depends upon the fixed constant $\varepsilon > 0$.

If we consider the disc with radius $\sqrt{N}$ and count the number of squares whose sum-of-digits function $s_q^{\mathbb{C}}(z^2)$ equals the "expected value" $\lfloor \mu_Q \log_Q |z|^4 \rfloor$, we get the following corollary:

Corollary 7.23. *Let $a \in \mathcal{A}$. We have, as $N \to \infty$,*

$$\#\left\{1 \leqslant |z| < \sqrt{N} : s_q^{\mathbb{C}}\left(z^2\right) = \lfloor \mu_Q \log_Q |z|^4 \rfloor\right\}$$
$$= \frac{\pi N}{(\log_Q N^2)^{\frac{1}{2}}} R\left(\frac{\mu_Q \log_Q N^2}{a^2+2a+2}\right) \cdot \left(1 + O\left(\frac{(\log\log N)^{11}}{(\log N)^{\frac{1}{2}}}\right)\right),$$

where $R(t)$ denotes a positive periodic function with period 1.

In order to show that Condition ($\mathbb{C}$3) holds true for the function $g(z) = z^2$, we have to deal with special exponential sum estimates. Gittenberger and Thuswaldner showed in [GT00, Chapter 2], that

$$\sum_{|z|^2<N} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{A}{B}z^2\right)\right) \ll N(\log N)^{-\sigma},$$

whenever $(A, B) = 1$ and $(\log N)^\sigma \ll |B|^2 \ll N^2(\log N)^{-\sigma}$. This estimate is too weak for us and it is only proved for discs with radius $\sqrt{N}$. In what follows, we use the van der Corput-type inequality Lemma 3.12 to treat such exponential sums. This enables us to improve the error term on the one hand and it allows us to consider 1/2-$\mathbb{Z}[i]$ sequences instead of discs on the other hand.

Lemma 7.24. *Let $A, B \in \mathbb{Z}[i]$ with $(A, B) = 1$ and let $(\mathcal{D}_N)_{N\in\mathbb{N}}$ be a 1/2-$\mathbb{Z}[i]$ sequence. Furthermore let $c > 0$ be real and $\sigma > 1$ such that*

$$Q^{4c(\log\log N)^\sigma} \ll |B|^2 \ll N^2 Q^{-4c(\log\log N)^\sigma}.$$

Then we have

$$\sum_{z\in\mathcal{D}_N} \mathrm{e}\left(\frac{1}{2}\operatorname{tr}\left(\frac{A}{B}z^2\right)\right) \ll NQ^{-c(\log\log N)^\sigma}(\log N),$$

where the implied constant is absolute.

Proof. Let S be the considered sum. Using Lemma 3.12 with $R = |B|^{1/2}$ yields

$$\begin{aligned} S &\ll \left(\frac{N}{|B|} \sum_{|r| \leqslant 2|B|^{1/2}} \left(1 - \frac{|r|}{2|B|^{1/2}+1} \right) \left| \sum_{z, z+r \in \Xi_N} \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{A}{B} \left((z+r)^2 - z^2 \right) \right) \right) \right| \right)^{1/2} \\ &\qquad + N^{1/2}|B|^{1/2} \\ &\ll \frac{N^{1/2}}{|B|^{1/2}} \left(\sum_{r \in R_B} \left| \sum_{z, z+r \in \Xi_N} \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{A}{B} 2rz \right) \right) \right| \right)^{1/2} + N^{1/2}|B|^{1/2}, \end{aligned}$$

where Ξ_N is defined at the beginning of Section 3.4.1. Hence we have to estimate linear exponential sums in $\mathbb{Z}[i]$. The important property is that we are summing over rectangles (with side length smaller than $2\sqrt{N}$). Thus we obtain

$$\left| \sum_{z, z+r \in \Xi_N} \mathrm{e}\left(\frac{1}{2} \operatorname{tr}\left(\frac{A}{B} 2rz \right) \right) \right| \ll \min\left(N, \frac{\sqrt{N}}{\|s_1\|}, \frac{\sqrt{N}}{\|s_2\|}, \frac{1}{\|s_1\| \cdot \|s_2\|} \right),$$

where $s_1 + is_2 = 2r\frac{A}{B}$. We therefore get

$$\begin{aligned} S \ll \frac{N^{1/2}}{|B|^{1/2}} &\left(\sum_{r \in R_B} \min\left(N, \frac{\sqrt{N}}{\|\Re\left(\frac{2A}{B}r\right)\|}, \frac{\sqrt{N}}{\|\Im\left(\frac{2A}{B}r\right)\|}, \frac{1}{\|\Re\left(\frac{2A}{B}r\right)\| \cdot \|\Im\left(\frac{2A}{B}r\right)\|} \right) \right)^{1/2} \\ &+ N^{1/2}|B|^{1/2}. \end{aligned}$$

If $(2A, B) = 1$, then $2Ar$ also runs through a complete residue system modulo B. Employing Lemma 3.14 yields

$$\begin{aligned} |S| &\ll \frac{N^{1/2}}{|B|^{1/2}} \left(N + |B| N^{1/2} \log|B| + |B|^2 (\log|B|)^2 \right)^{1/2} + N^{1/2}|B|^{1/2} \\ &\ll (\log|B|) \left(N|B|^{-1/2} + N^{1/2}|B|^{1/2} \right). \end{aligned}$$

Using the bounds on $|B|^2$ brings the desired result. If $(2A, B) \neq 1$ we write $\frac{2A}{B} = \frac{A'}{B'}$ with $(A', B') = 1$ and we can do similar calculations as above. Since B and B' are comparable (note that $(A, B) = 1$) we obtain the same result in this case, too. □

Proof of Theorem 7.21. A κ-$\mathbb{Z}[i]$ sequence satisfies the desired properties stated in Theorem 7.3. Thus, we only have to verify Conditions ($\mathbb{C}$1)–($\mathbb{C}$3) for the function $g(z) = z^2$. We know from Proposition A.5 that there exists a constant $\xi \geqslant 0$, such that the Gaussian integer z^2 has less than or equal to $\log_Q N^2 + \xi$ digits whenever $z \in \mathcal{D}_N$. Lemma 3.17 implies that

$$\frac{1}{\#\mathcal{D}_N} \#\{z \in \mathcal{D}_N : z^2 \equiv k \bmod q-1\} = Q(k, q-1) + O\left(N^{-1/2}\right).$$

Hence Condition ($\mathbb{C}$1) holds true. Condition ($\mathbb{C}$2) also holds true (see Theorem 3.2) and Condition ($\mathbb{C}$3) follows from the previous lemma. □

Proof of Corollary 7.23. Corollary 7.23 follows immediately from Theorem 7.21. The proof is analog to the proofs of Lemma 7.17 and [DMR09, Theorem 1.2]. Due to the fact that we are dealing with Gaussian squares we carry out all necessary steps. The most important property is that an integer $m \in \mathbb{Z}$ satisfies $m = \lfloor \mu_Q \log_Q |z|^4 \rfloor$ if and only if $Q^{m/2\mu_Q} \leqslant |z|^2 < Q^{(m+1)/2\mu_Q}$. We define $A_m(N) = \#\{|z| < \sqrt{N} : s_q^{\mathbb{C}}(z^2) = m\}$ and set $\widetilde{\ell} = \left\lfloor \frac{\mu_Q \log_Q N^2}{a^2+2a+2} \right\rfloor$. This allows us to write

$$\#\{|z| < \sqrt{N} : s_q^{\mathbb{C}}(z^2) = \lfloor \mu_Q \log_Q |z|^4 \rfloor\} = S_1 + S_2 + S_3,$$

where

$$S_1 = \sum_{m < \widetilde{\ell}(a^2+2a+2)} \left(A_m(Q^{(m+1)/2\mu_Q}) - A_m(Q^{m/2\mu_Q}) \right),$$

$$S_2 = \sum_{m=\widetilde{\ell}(a^2+2a+2)}^{\lfloor \mu_Q \log_Q N^2 \rfloor - 1} \left(A_m(Q^{(m+1)/2\mu_Q}) - A_m(Q^{m/2\mu_Q}) \right),$$

and

$$S_3 = A_{\lfloor \mu_Q \log_Q N^2 \rfloor}(N) - A_{\lfloor \mu_Q \log_Q N^2 \rfloor}(Q^{\lfloor \mu_Q \log_Q N^2 \rfloor / 2\mu_Q}).$$

Theorem 7.21 implies that for $j \in \{0, 1\}$ we have

$$A_m(Q^{\frac{m+j}{2\mu_Q}}) = \frac{Q(m, q-1)}{\sqrt{2\pi\sigma_Q^2}} \frac{Q^{\frac{m+j}{2\mu_Q}} \pi}{(m/\mu_Q)^{\frac{1}{2}}} \left(1 + O\left(\frac{(\log m)^{11}}{m^{\frac{1}{2}}} \right) \right).$$

(Note that this is also true if $Q(m, q-1) = 0$.) Since $Q(\,.\,, q-1)$ is periodic with period $a^2 + 2a + 2$, we get

$$\begin{aligned} S_1 = {} & \frac{\pi}{\sqrt{2\pi\sigma_Q^2}} \sum_{1 \leqslant \ell < \widetilde{\ell}} \frac{Q^{\frac{\ell(a^2+2a+2)}{2\mu_Q}}}{\sqrt{\frac{\ell(a^2+2a+2)}{\mu_Q}}} \left(Q^{\frac{1}{2\mu_Q}} - 1 \right) \\ & \cdot \sum_{0 \leqslant k < a^2+2a+2} Q(k, q-1) Q^{\frac{k}{2\mu_Q}} \left(1 + O\left((\log \ell)^{11} \ell^{-\frac{1}{2}} \right) \right). \end{aligned}$$

Setting

$$c_1 = \frac{Q^{\frac{1}{2\mu_Q}} - 1}{Q^{\frac{a^2+2a+2}{2\mu_Q}} - 1} \cdot \sum_{0 \leqslant k < a^2+2a+2} Q(k, q-1) Q^{\frac{k}{2\mu_Q}},$$

we obtain by using partial summation

$$\begin{aligned} S_1 &= \frac{\pi c_1}{\sqrt{2\pi\sigma_Q^2}} \frac{Q^{\frac{\widetilde{\ell}(a^2+2a+2)}{2\mu_Q}}}{\sqrt{\frac{\widetilde{\ell}(a^2+2a+2)}{\mu_Q}}} \left(1 + O\left((\log \widetilde{\ell})^{11} \widetilde{\ell}^{-\frac{1}{2}} \right) \right) \\ &= \frac{c_1}{\sqrt{2\pi\sigma_Q^2}} \frac{\pi N}{\sqrt{\log_Q N^2}} Q^{-\left\{ \frac{\mu_Q \log_Q N^2}{a^2+2a+2} \right\} \frac{a^2+2a+2}{2\mu_Q}} \left(1 + O\left((\log\log N)^{11} (\log N)^{-\frac{1}{2}} \right) \right). \end{aligned}$$

To calculate S_2, we set

$$c_2(t)=\left(Q^{\frac{1}{2\mu_Q}}-1\right)\sum_{0\leqslant k<t(a^2+2a+2)-\{t(a^2+2a+2)\}}Q(k,q-1)q^{\frac{k}{2\mu_Q}}.$$

Then we get

$$S_2=\frac{c_2\left(\left\{\frac{\mu_Q\log_Q N^2}{a^2+2a+2}\right\}\right)}{\sqrt{2\pi\sigma_Q^2}}\frac{\pi N}{\sqrt{\log_Q N^2}}Q^{-\left\{\frac{\mu_Q\log_Q N^2}{a^2+2a+2}\right\}\frac{a^2+2a+2}{2\mu_Q}}$$
$$\cdot\left(1+O\left((\log\log N)^{11}(\log N)^{-\frac{1}{2}}\right)\right).$$

Similar we have

$$S_3=\frac{Q(\lfloor\mu_Q\log_Q N^2\rfloor,q-1)}{\sqrt{2\pi\sigma_Q^2}}\frac{\pi N}{\sqrt{\log_Q N^2}}\left(1-Q^{-\frac{\{\mu_Q\log_Q N^2\}}{2\mu_Q}}\right)$$
$$\cdot\left(1+O\left((\log\log N)^{11}(\log N)^{-\frac{1}{2}}\right)\right).$$

Note, that $\lfloor\mu_Q\log_Q N^2\rfloor\equiv\left\lfloor\left\{\frac{\mu_Q\log_Q N^2}{a^2+2a+2}\right\}(a^2+2a+2)\right\rfloor \bmod (a^2+2a+2)$. Defining the function $R(t)$ by

$$R(t)=\frac{1}{\sqrt{2\pi\sigma_Q^2}}\Bigg((c_1+c_2(\{t\}))Q^{-\{t\}\frac{a^2+2a+2}{2\mu_Q}}$$
$$+Q\Big(\lfloor\{t\}(a^2+2a+2)\rfloor,q-1\Big)\left(1-Q^{-\frac{\{t(a^2+2a+2)\}}{2\mu_Q}}\right)\Bigg),$$

implies the desired result. □

Appendix

In this appendix we collect important and well-known results, which we used throughout this thesis. In particular, we give a short overview of the base-q representation in the ring of Gaussian integers, we state Fourier transform estimates of the sum-of-digits function in the natural numbers and we cite classical Gauss sums. Moreover, we state some established exponential sum estimates and consider approximation properties of the Beurling-Selberg function.

A.1 Digital expansion in the ring of Gaussian integers

Recall that every positive integer has a unique expansion in base $q \geqslant 2$. A natural extension of the q-ary representation system to the ring of Gaussian integers provides the following definition (see for example [GT00, Definition 1.1]):

Definition A.1. A pair $(q, \mathcal{N})$ with $q \in \mathbb{Z}[i]$ and $\mathcal{N} = \{0, 1, \dots, |q|^2 - 1\}$ is called canonical number system if every $z \in \mathbb{Z}[i]$ has a representation of the form

$$z = \varepsilon_0(z) + \varepsilon_1(z)q + \cdots + \varepsilon_{\ell-1}q^{\ell-1}$$

with $\ell \in \mathbb{Z}^+$ and $\varepsilon_j(z) \in \mathcal{N}$ for all $0 \leqslant j < \ell$. The Gaussian integer q is called the base and $\mathcal{N}$ is called the set of digits.

Since $\mathcal{N}$ is a complete residue system modulo q, the representation is unique. Kátai and Szabó [KS75] characterized all posible canonical number systems (also called base-q representation systems) in the Gaussian integers.

Theorem A.2. *The only canonical number systems are given by the bases $q = -a+i$ and $q = -a-i$ with $a \geqslant 1$.*

In order to find the digits of a given Gaussian integer z, Gilbert [Gil] described two methods. The *Standard Base Conversion Algorithm* works just as in the real case (chain of divisions) and the *Clearing Algorithm* uses the minimal polynom of $q = -a \pm i$. Based on this clearing algorithm, Grabner, Kirschenhofer and Prodinger gave in [GKP98, Proposition 2.2, Corollary 2.3] an explicit recursion for the digits which allowed them to prove an interesting representation of the sum-of-digits function.

Lemma A.3. *Let $q = -a \pm i$. For $z = z_1 + iz_2 \in \mathbb{Z}[i]$ let $\sigma_k(z) \in \mathbb{Z}$ be defined by the recurrence (the choice of the initial values depends on the choice of the bases $q = -a \pm i$)*

$$\sigma_{k+1}(z) = (a-1)^2 \left\lfloor \frac{\sigma_k(z)}{a^2+1} \right\rfloor + (2a-1) \left\lfloor \frac{\sigma_{k-1}(z)}{a^2+1} \right\rfloor + \left\lfloor \frac{\sigma_{k-2}(z)}{a^2+1} \right\rfloor,$$

with

$$\sigma_{-2}(z) = \pm(a^2+1)z_2, \qquad \sigma_{-1}(z) = 0, \qquad \sigma_0(z) = z_1 \pm az_2.$$

Then we have $z = \sum_{j \geqslant 0} \varepsilon_j(z) q^j$ with

$$\varepsilon_j(z) = \sigma_j(z) \bmod (a^2+1).$$

The sum-of-digits function is defined in the following way: Recall that if $z \in \mathbb{Z}[i]$, then it has a unique expansion of the form $z = \sum_{j=0}^{\ell-1} \varepsilon_j(z) q^j$, $\varepsilon_j(z) \in \mathcal{N}$. The (complex) sum-of-digits function $s_q^{\mathbb{C}}$ is given by $z \mapsto \sum_{j \geqslant 0} \varepsilon_j(z)$

Proposition A.4. *Let $q = -a \pm i$. For $z = z_1 + iz_2 \in \mathbb{Z}[i]$ the sequence $(\sigma_k(z))$ from Lemma A.3 is ultimately constant. Denoting the limit by $\sigma_\infty(z)$, this value is divisible by a^2+1, and the sum-of-digits function satisfies*

$$s_q^{\mathbb{C}}(z) = z_1 \pm (a+1)z_2 - (a^2+2a+2)\frac{\sigma_\infty(z)}{a^2+1},$$

where the choice of the sign depends on the choice of the sign in $q = -a \pm i$.

The following lemma provides information about the number of digits of a given Gaussian integer z. Roughly speaking, z has $\log_Q(|z|^2) + O(1)$ digits in its base-q representation system. A proof can be found in [GKP98, Proposition 2.6].

Proposition A.5. *Let ℓ be the smallest number satisfying $z = \sum_{0 \leqslant j < \ell} \varepsilon_j(z) q^j$ with $\varepsilon_j(z) \in \mathcal{N}$. Then we have (setting $Q = |q|^2$)*

$$\log_Q |z|^2 - 2\log_Q \frac{a\sqrt{a^2+4}}{a^2+2} - 4 \leqslant \ell \leqslant \log_Q |z|^2 - \log_Q \left(1 - \frac{a\sqrt{a^2+4}}{a^2+2}\right) + 4.$$

Next, we give a short overview on the radix expansion of complex numbers. Kátai and Szabó [KS75] not only characterized the valid bases for representing Gaussian integers, but also proved that for $q = -a \pm i$, $a \in \mathbb{Z}^+$ every complex number z can be written in the form

$$z = \varepsilon_\ell q^\ell + \cdots + \varepsilon_0 + \frac{\varepsilon_{-1}}{q} + \frac{\varepsilon_{-2}}{q^2} + \cdots,$$

where $\varepsilon_j \in \mathcal{N}$. This (possibly) infinite expansion is denoted by

$$\varepsilon_\ell\, \varepsilon_{\ell-1} \cdots \varepsilon_0 \,.\, \varepsilon_{-1}\, \varepsilon_{-2} \cdots$$

and the number left to the radix point is called the integer part of the expansion. The set of complex numbers with integer part zero

$$\mathcal{F}' = \left\{ z \in \mathbb{C} : z = \sum_{j=1}^{\infty} \epsilon_j(z) q^{-j}, \varepsilon_j \in \mathcal{N} \right\}$$

is called the fundamental domain of the base-q representation system. It is a compact set of unit area and it forms a tiling of the complex plain. The boundary of $\mathcal{F}'$ in any base $q = -a \pm i$ has Hausdorff dimension > 1 and thus is a fractal. Points on it are complex numbers that have (at least) two representations with different integer parts (one of them being zero). Interestingly, if $q \neq -2 \pm i$, then there are precisely six numbers which have three different radix expansions. For example, we have in base $q = -1 + i$ that

$$(1 - 2i)/5 = 0.001001001\ldots = 1.100100100\ldots = 111.010010010\ldots,$$

where the expansions are all periodic of period three. In the case $q = -2 \pm i$ the fundamental domain has a much more jagged reentrant form and there are a countable number of such points. In any case, no complex number has four different radix expansions. Another interesting fact is that $\mathcal{F}'$ is arcwise connected and that every point in $\mathcal{F}'$ with finite base-q representation is an inner point of $\mathcal{F}'$. For these results and the connection of fractal geometry and complex bases see for example Gilbert [Gil82, Gil87], Akiyama and Thuswaldner [AT00] and Müller, Thuswaldner and Tichy [MTT01].

A.2 The Fourier transform of the sum-of-digits function

Let $q \geqslant 2$, $\alpha \in \mathbb{R}$ and $\lambda \in \mathbb{N}$. The discrete Fourier transform $F_\lambda(., \alpha)$ of the function $u \mapsto \mathrm{e}(\alpha s_q(u))$, $0 \leqslant u < q^\lambda$ is defined for all $h \in \mathbb{Z}$ by

$$F_\lambda(h, \alpha) = \frac{1}{q^\lambda} \sum_{0 \leqslant u < q^\lambda} \mathrm{e}\left(\alpha s_q(u) - huq^{-\lambda}\right),$$

where $\mathrm{e}(x) = e^{2\pi i x}$ and s_q denotes the sum-of-digits function in $\mathbb{N}$. This function is periodic with period q^λ in the first component and can be represented by a trigonometric product. In particular, we have for the absolute value

$$|F_\lambda(h, \alpha)| = q^{-\lambda} \prod_{1 \leqslant j \leqslant \lambda} \varphi_q\left(\alpha - hq^{-j}\right),$$

where φ_q is defined by

$$\varphi_q(t) = \begin{cases} \frac{|\sin \pi q t|}{|\sin \pi t|}, & \text{if } t \in \mathbb{R} \setminus \mathbb{Z}, \\ q, & \text{if } t \in \mathbb{Z}. \end{cases}$$

For a thorough analysis of φ_q and F_λ see [MR10, MR09] and [FM96]. The following lemma collects some basic facts of $\varphi_k(t)$. Proofs can be found in [MR10, Lemmas 14, 15] and [MR09, Lemmas 3, 5].

Lemma A.6. *Let k be an integer $\geqslant 2$. Then, the following claims hold true:*

(i) *The function $\varphi_k(t)$ is periodic of period 1, monotonically decreasing on $[0, 1/k]$ and we have for $\delta \in [0, 2/(3k)]$,*

$$\max_{\|t\| \geqslant \delta} \varphi_k(t) \leqslant \varphi_k(\delta) \leqslant k.$$

(ii) *If $\|t\| \leqslant \sqrt{\frac{6}{\pi^2(k^2-1)}}$, then we have*

$$\varphi_k(t) \leqslant k \exp\left(-\frac{(k^2-1)\pi^2 \|t\|^2}{6} \right).$$

(iii) *Let $\Phi(k)$ be defined by $\Phi(k) = \max_{t \in \mathbb{R}} \left(\frac{1}{k} \sum_{0 \leqslant r < k} \varphi_k \left(t + \frac{r}{k} \right) \right)$. Then*

$$\Phi(k) \leqslant \frac{2}{k \sin \frac{\pi}{2k}} + \frac{2}{\pi} \log \frac{2k}{\pi}.$$

Assume that $k \geqslant 3$ and that η_k is given by $k^{\eta_k} = \Phi(k)$. Then we have $0.24998 < \eta_{689} < 0.24999$ and $\eta_k < \eta_{689}$ for $k > 689$.

(iv) *Let $k \geqslant 3$. If $3 \leqslant R \leqslant k$ and $R \mid k$, then*

$$\max_{t \in \mathbb{R}} \left(\frac{1}{k} \sum_{0 \leqslant r < R} \varphi_k \left(t + \frac{r}{R} \right) \right) \leqslant R^{\eta_R}.$$

If $R = 2$ and $R \mid k$, then the right-hand side of the given estimate can be replaced by $\sqrt{3/2} < 1.23 < 1.34 < 2^{\eta_5}$.

(v) *We have for all $t \in \mathbb{R}$ the inequality*

$$\varphi_k \left(\frac{t}{k} \right) \leqslant \varphi_k \left(\frac{\|t\|}{k} \right).$$

The following lemma deals with an L^∞-norm of F_λ and can be found in [MR09, Lemma 9].

Lemma A.7. *Let $q \geqslant 2$, $\alpha \in \mathbb{R}$, $h \in \mathbb{Z}$, $\lambda \geqslant 1$ and $\sigma_q = \frac{\pi^2}{12 \log q} \left(1 - \frac{2}{q+1} \right)$. Then we have*

$$|F_\lambda(h, \alpha)| \leqslant e^{\pi^2/48} q^{-\sigma_q \|(q-1)\alpha\|^2 \lambda}.$$

We distinguish two different cases when considering the L^1-norm of the Fourier transform $\mathcal{F}_\lambda$. The first one gives an upper bound of the norm if $q \geqslant 3$ (η_q is defined in Lemma A.6), see [MR10, Lemmas 16 and 17].

Lemma A.8. *For $q \geqslant 3$, $\alpha \in \mathbb{R}$, $a \in \mathbb{Z}$, $0 \leqslant \delta \leqslant \lambda$, $k \mid q^{\lambda-\delta}$ and $k \nmid q$, we have*

$$\sum_{\substack{0 \leqslant h < q^\lambda \\ h \equiv a \bmod kq^\delta}} |F_\lambda(h,\alpha)| \leqslant k^{-\eta_3} q^{\eta_3(\lambda-\delta)} |F_\delta(a,\alpha)|.$$

Moreover, we have

$$\sum_{0 \leqslant h < q^\lambda} |F_\lambda(h,\alpha)| \leqslant q^{\eta_q \lambda}.$$

Remark A.9. The Cauchy-Schwarz inequality together with Parseval's identity implies that

$$\sum_{0 \leqslant h < q^\lambda} |F_\lambda(h,\alpha)| \leqslant q^{\eta_q \lambda}$$

holds true for sure with $\eta_q \leqslant 1/2$. Interestingly, the lemma stated above gives a much better estimate if q is big. In particular, part (iii) of Lemma A.6 implies that $\eta_q \leqslant \log\log q / \log q$ for $q \geqslant 15$ and hence η_q is arbitrary small if q is big enough.

For the sake of completeness, we also state the L^1-norm estimate in the case of $q = 2$ (see [MR10, Lemma 18]):

Lemma A.10. *For $q = 2$ we define η_2 by the equation*

$$2^{\eta_2} = (2+\sqrt{2})^{1/4} \quad (\textit{in particular } 0.4428 < \eta_2 < 0.4429).$$

Then we have for all $\alpha \in \mathbb{R}$, $a \in \mathbb{Z}$ and $0 \leqslant \delta \leqslant \lambda$

$$\sum_{\substack{0 \leqslant h < 2^\lambda \\ h \equiv a \bmod 2^\delta}} |F_\lambda(h,\alpha)| \leqslant 2^{\eta_2(\lambda-\delta)+1/2} |F_\delta(a,\alpha)|.$$

A.3 Gauss sums

Gauss introduced in 1801 (Disquisitiones Arithmeticae) the sum $\sum_{n=0}^{c-1} \mathrm{e}\left(an^2/c\right)$. Until nowadays, a lot of different generalizations have been introduced (see for example [BEW98, GK91] and [IK04, Chapter 3.4]). It is well studied in the literature and in particular, if $(c,a) = 1$, we have

$$\sum_{n=0}^{c-1} \mathrm{e}\left(\frac{an^2}{c}\right) = \begin{cases} \left(\frac{c}{a}\right)(1+i^a)\sqrt{c}, & \text{if } c \equiv 0 \bmod 4, \\ \left(\frac{a}{c}\right)\sqrt{c}, & \text{if } c \equiv 1 \bmod 4, \\ 0, & \text{if } c \equiv 2 \bmod 4, \\ i\left(\frac{a}{c}\right)\sqrt{c}, & \text{if } c \equiv 3 \bmod 4, \end{cases}$$

where $\left(\frac{r}{s}\right)$ denotes the Jacobi symbol (see for example [BEW98, Chapter 1]). If we set more generally

$$G(a,\ell;c) = \sum_{n=0}^{c-1} \mathrm{e}\left(\frac{an^2+\ell n}{c}\right).$$

then the following proposition holds true (this is a consequence of [GK91, Inequality (7.4.2)]):

Proposition A.11. *Let $a, \ell, c \in \mathbb{Z}$ with $c \geqslant 1$ and set $d = (a, c)$. Then we have*

$$|G(a, \ell; c)| \leqslant \sqrt{2dc},$$

and

$$|G(a, \ell; c)| = 0 \quad \textit{if } d \nmid \ell.$$

A.4 Exponential sum estimates

Exponential sums, that is, sums of the form

$$\sum_{a<n\leqslant b} \mathrm{e}(f(n)),$$

play an important role in analytic number theory. Here, and throughout this section, f is a real-valued function and $\mathrm{e}(x) = e^{2\pi i x}$. Since exponential sums are in general hard to compute, one is typically interested in estimates of such sums. These estimates mostly depend on the length of the summation interval and on the growth of the derivatives of the function f. We denote the summation interval $(a, b]$ by I and the length of it by $|I|$. In what follows, we state several well-known results which use different assumptions on f. For a thorough and up-to-date treatment of exponential sums see for example Graham and Kolesnik [GK91] and Iwaniec and Kowalski [IK04, Chapter 8].

The first theorem (see [GK91, Theorem 2.1]) is named after Kusmin and Landau and just uses the first derivative of f ($\|\cdot\|$ denotes the distance to the nearest integer):

Theorem A.12. *If f is continuously differentiable, f' is monotonic, and $\|f'\| \geqslant \lambda > 0$ on I, then*

$$\sum_{n\in I} \mathrm{e}(f(n)) \ll \lambda^{-1}.$$

If f is linear, then this result is sharp. The following theorem (attributed to van der Corput) also uses the second derivative and gives much better results if f is “far away” from being linear (see [GK91, Theorem 2.2]). The idea of the proof consists of replacing the sum by integrals and using the van der Corput inequality in order to obtain the desired estimate.

Theorem A.13. *Suppose that f has two continuous derivatives on I. Suppose also that there is some $\lambda > 0$ and some $\alpha \geqslant 1$ such that*

$$\lambda \leqslant \left|f''(x)\right| \leqslant \alpha\lambda$$

on I. Then we have

$$\sum_{n\in I} \mathrm{e}(f(n)) \ll \alpha|I|\lambda^{1/2} + \lambda^{-1/2}.$$

The idea of using ever higher derivatives is contained in the following theorem (see [GK91, Theorem 2.9]):

Theorem A.14. *Let $q \geqslant 0$ be an integer. Suppose that f has $q+2$ continuous derivatives on I and that $I \subseteq (N, 2N]$. Assume also that there is some constant F such that*

$$FN^{-r} \ll |f^{(r)}(x)| \ll FN^{-r} \tag{A.24}$$

for $r = 1, \dots, q+2$ on I. Let $Q = 2^q$. Then we have

$$\sum_{n \in I} \mathrm{e}(f(n)) \ll F^{1/(4Q-2)} N^{1-(q+2)/(4Q-2)} + F^{-1}N.$$

The implied constant depends only upon the implied constants in (A.24).

The results presented so long are obtained by using the so-called van der Corput method of exponential sums. They give rise to the method of exponential pairs which we do not treat here but which is of particular interest (see for example [GK91, Chapter 3] and [IK04, Section 8.4]). The next two theorems can be proved by Vinogradov's method, which is very strong for sums of large amplitude relative to the length of the summation interval. Theorem A.15 is due to Vinogradov himself (see [Vin04, Theorem 2a, p. 109]), Theorem A.16 is an improvement for large values of the amplitude of Vinogradov's theorem (see [IK04, Theorem 8.25]).

Theorem A.15. *Suppose that f has a $(n+1)$-th continuous derivative on I and that $I \subseteq (N, 2N]$. Assume also that there is a constant F such that*

$$F \leqslant \frac{x^{n+1}}{(n+1)!} \left| f^{(n+1)}(x) \right| \ll_n F,$$

and

$$N \leqslant F^{-1} N^{n+1} \leqslant N^{2+2/n}.$$

Then we have

$$\sum_{n \in I} \mathrm{e}(f(n)) \ll N^{1-1/(3n^2 \log(125n))}.$$

Theorem A.16. *Let f be a smooth function on $(N, 2N]$ which satisfies for all $j \geqslant 1$*

$$\alpha^{-j^3} F \leqslant \frac{x^j}{j!} \left| f^{(j)}(x) \right| \leqslant \alpha^{j^3} F$$

on $(N, 2N]$, where $F \geqslant N^4$ and $\alpha \geqslant 1$. Let $I \subseteq (N, 2N]$. Then we have

$$\sum_{n \in I} \mathrm{e}(f(n)) \ll \alpha N \exp(-2^{-18} (\log N)^3 (\log F)^{-2}),$$

where the implied constant is absolute.

Although the following theorem does not fit perfectly to the previous results, it can be used ot estimate certain sums, which are often related to exponential sums. It is due to Bombieri and Iwaniec (see [BI86, Lemma 2.4]) and it is called a "double large sieve".

Theorem A.17. *Let $\mathcal{X}$ and $\mathcal{Y}$ be two sets in $\mathbb{R}^\kappa$, $\kappa \geqslant 1$ and let $\alpha(x)$ for $x = (x_1, \dots, x_\kappa) \in \mathcal{X}$ and $\beta(y)$ for $y = (y_1, \dots, y_\kappa) \in \mathcal{Y}$ be arbitrary complex numbers. Furthermore, let $X_1, \dots, X_\kappa$ and $Y_1, \dots, Y_\kappa$ be positive numbers and set*

$$\mathcal{B}_1(\alpha) = \sum_{\substack{x,x' \in \mathcal{X} \\ |x_k - x'_k| \leqslant (2Y_k)^{-1} \\ k=1,\dots,\kappa}} \left|\alpha(x)\alpha(x')\right|, \qquad \mathcal{B}_2(\beta) = \sum_{\substack{y,y' \in \mathcal{Y} \\ |y_k - y'_k| \leqslant (2X_k)^{-1} \\ k=1,\dots,\kappa}} \left|\beta(y)\beta(y')\right|,$$

and

$$\mathcal{B}(a,b) = \sum_{\substack{x \in \mathcal{X},\, y \in \mathcal{Y} \\ |x_k| \leqslant X_k,\, |y_k| \leqslant Y_k \\ k=1,\dots,\kappa}} \alpha(x)\beta(y)\,\mathrm{e}(x \cdot y).$$

Then we have

$$|\mathcal{B}(\alpha,\beta)|^2 \leqslant \frac{1}{(2\pi^2)^\kappa} \prod_{k=1}^{\kappa} (1 + X_k Y_k) \mathcal{B}_1(\alpha) \mathcal{B}_2(\beta).$$

In the last part we shortly discuss the theory of uniform distribution modulo 1. As we will see, exponential sum estimates play an important role in this field of study. Let $\mathbf{x} = (x_n)_{n \in \mathbb{N}}$ be a sequence of points in the k-dimensional space $\mathbb{R}^k$. If $I = [a_1, b_1) \times \cdots \times [a_k, b_k)$ is an interval on the torus $\mathbb{R}^k/\mathbb{Z}^k$, then $A(I, N, \mathbf{x})$ denotes the number of points x_n, $0 \leqslant n \leqslant N-1$, for which $\{x_n\} \in I$, where the fractional part $\{x\} = x - \lfloor x \rfloor$ is taken componentwise. The sequence $\mathbf{x} = (x_n)_{n \in \mathbb{N}}$ is said to be uniformly distributed modulo 1 if for every interval $I \subseteq \mathbb{R}^k/\mathbb{Z}^k$ we have

$$\lim_{N \to \infty} \frac{A(I, N, \mathbf{x})}{N} = \lambda_k(I),$$

where $\lambda_k(I)$ denotes the k-dimensional Lebesgue measure of I. A sequence $\mathbf{x} = (x_n)_{n \in \mathbb{N}}$ is uniformly distributed modulo 1 if and only if the discrepancy

$$D_N(\mathbf{x}) = \sup_{I \subseteq \mathbb{R}^k/\mathbb{Z}^k} \left| \frac{A(I, N, \mathbf{x})}{N} - \lambda_k(I) \right|$$

satisfies $D_N(\mathbf{x}) = o(1)$. The next theorem is due to Weyl and establishes a connection between the theory of uniform distribution modulo 1 and the theory of exponential sum estimates (see for example [DT97, Theorem 1.19]).

Theorem A.18 (Weyl's criterion)**.** *A sequence $\mathbf{x} = (x_n)_{n \in \mathbb{N}}$ of points in the k-dimensional space $\mathbb{R}^k$ is uniformly distributed modulo 1 if and only if*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{0 \leqslant n < N} \mathrm{e}(h \cdot x_n) = 0$$

holds for all non-zero integral lattice points $h \in \mathbb{Z}^k \setminus \{(0, \dots, 0)\}$.

Finally, we state a result that shows how exponential sums can be used to estimate the discrepancy.

Theorem A.19 (Erdős-Turán-Koksma inequality). *Let* $\mathbf{x} = (x_n)_{n\in\mathbb{N}}$ *be a sequence of points in the* k*-dimensional space* $\mathbb{R}^k$ *and* H *an arbitrary positive integer. Then we have*

$$D_N(\mathbf{x}) \ll_k \frac{1}{H+1} + \sum_{0<\|h\|_\infty \leqslant H} \frac{1}{r(h)} \left| \frac{1}{N} \sum_{0\leqslant n<N} \mathrm{e}(h\cdot x_n) \right|,$$

where $r(h) = \prod_{1\leqslant i\leqslant k} \max(1, |h_i|)$ *for* $h = (h_1, \ldots, h_k) \in \mathbb{Z}^k$.

There are some admissible values known for the implied constant in the last theorem (see [DT97, Theorem 1.21]). For $k = 1$ (in this case the above result is called Erdős-Turán inequality) Mauduit, Rivat, and Sárközy [MRS02, Lemma 1] showed that

$$D_N(\mathbf{x}) \leqslant \frac{1}{H+1} + \sum_{h=1}^{H} \frac{1}{h} \left| \frac{1}{N} \sum_{0\leqslant n<N} e\left(\frac{h}{x_n}\right) \right|. \tag{A.25}$$

For a further discussion of the implied constant in dimension 1 see [RT05].

A.5 Extremal functions in Fourier analysis

In the late 1930's Beurling observed that the entire function of exponential type 2π

$$B(z) = \left(\frac{\sin \pi z}{\pi}\right)^2 \left\{ \sum_{n=0}^{\infty} (z-n)^2 - \sum_{m=1}^{\infty} (z+m)^2 + 2z^{-1} \right\}$$

satisfies a simple and useful extremal property. If $\operatorname{sgn}(x)$ denotes the sign of x (with the usual convention $\operatorname{sgn}(0) = 0$), then

$$\operatorname{sgn}(x) \leqslant B(x)$$

for real x and

$$\int_{-\infty}^{\infty} B(x) - \operatorname{sgn}(x)\,\mathrm{d}x = 1.$$

Moreover, he showed that if $F(z)$ is any entire function of exponential type 2π satisfying $\operatorname{sgn}(x) \leqslant F(x)$ for real x, then $\int_{-\infty}^{\infty} F(x) - \operatorname{sgn}(x)\,\mathrm{d}x \geqslant 1$ where equality holds if and only if $F(z) = B(z)$. In 1974, Selberg used this function to obtain a sharp form of the large sieve inequality. Nowadays, the function $B(z)$ is therefore often called Beurling-Selberg function. In 1985, Vaaler used $B(z)$ in order to approximate special functions with trigonometric polynomials. In the following, we present some of his results (taken from [Vaa85]). For many other interesting properties of the

Beurling-Selberg function and related functions we refer to the original work [Vaa85] and to [GK91, Appendix] for a short discussion of Vaaler's results. Set

$$\psi^\circ(x) = \begin{cases} x - \lfloor x \rfloor - 1/2, & \text{if } x \notin \mathbb{Z}, \\ 0, & \text{if } x \in \mathbb{Z}, \end{cases}$$

and

$$J(z) = \frac{1}{2} \cdot \frac{\mathrm{d}}{\mathrm{d}z} \left\{ B(z) - \left(\frac{\sin \pi z}{\pi z} \right)^2 \right\}.$$

The function $J(z)$ has the remarkable property that its Fourier transform $\hat{J}(t)$ has compact support. In particular, it is given by

$$\hat{J}(t) = \begin{cases} 1, & \text{if } t = 0, \\ \pi t(1 - |t|) \cot \pi t + |t|, & \text{if } 0 < |t| < 1, \\ 0, & \text{if } |t| \geqslant 1. \end{cases}$$

Set $J_N(z) = NJ(Nz)$ and let $j_N(x)$ be defined by $j_N(x) = \sum_{m=-\infty}^{\infty} J_{N+1}(x+m)$. Using Poisson's summation formula we get

$$j_N(x) = \sum_{n=-N}^{N} \hat{J}_{N+1}(n)\, \mathrm{e}(nx), \tag{A.26}$$

where $\hat{J}_N(t) = \hat{J}(t/N)$. Finally, let $\kappa_N(x)$ be defined by

$$\kappa_N(x) = \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N+1} \right) \mathrm{e}(nx).$$

This function is known as the the periodic Fejer kernel and it is positive for real numbers x. (This can be easily seen since $\kappa_N(x)$ is also given by $\kappa_N(x) = (N+1)^{-1}(\sin \pi(N+1)x)^2/(\sin \pi x)^2$.) Vaaler showed among other things the following connection between these functions (see [Vaa85, Theorem 18]):

Proposition A.20. *The trigonometric polynomial*

$$\psi_N(x) = -\frac{1}{2i\pi} \sum_{1 \leqslant |n| \leqslant N} \frac{1}{n} \hat{J}_{N+1}(n)\, \mathrm{e}(nx)$$

satisfies

$$|\psi_N(x) - \psi^\circ(x)| \leqslant \frac{1}{2N+2} \kappa_N(x).$$

We want to remark that by continuity this result holds also true if $\psi^\circ(x)$ is replaced by $\psi(x) = x - \lfloor x \rfloor - 1/2$. The following proposition is a generalization of this result, see [Vaa85, Theorem 19]. Let f be a function of period 1 which has bounded variation on each closed interval of length 1. Suppose furthermore that f satisfies the normalizing condition

$$\lim_{h\to 0+} \frac{1}{2}(f(x+h)+f(x-h)) = f(x) \tag{A.27}$$

for every $x \in \mathbb{R}$. Denote by $V_f(x)$ the total variation of f on $[-1/2, x]$ and let $(\mathrm{d}V_f) * \kappa_N(x)$ be the convolution

$$(\mathrm{d}V_f) * \kappa_N(x) = \int_{-1/2}^{1/2} \kappa_N(x-\xi)\mathrm{d}V_f(\xi).$$

Furthermore, if f and g are two functions, let $f * g(x)$ denote the convolution

$$f * g(x) = \int_{-1/2}^{1/2} f(x-\xi)g(\xi)\mathrm{d}\xi.$$

Proposition A.21. *The trigonometric polynomial* $f * j_N(x)$ *and* $(\mathrm{d}V_f) * \kappa_N(x)$ *satisfy*

$$|f(x) - f * j_N(x)| \leqslant \frac{1}{2N+2}\,(\mathrm{d}V_f) * \kappa_N(x).$$

Equation (A.26) shows that the function $\psi_N(x)$ of Proposition A.20 is given by the convolution $\psi_N(x) = \psi * j_N(x)$. More generally, if $\hat{f}(n)$ denotes the Fourier coefficient

$$\hat{f}(n) = \int_{-1/2}^{1/2} f(x)\,\mathrm{e}(-nx)\mathrm{d}x,$$

then $f * j_N(x)$ is given by

$$f * j_N(x) = \sum_{n=-N}^{N} \hat{f}(n)\hat{J}_{N+1}(n)\,\mathrm{e}(nx).$$

Take for example $\chi(t) = \mathbf{1}_A(\{t\})$, where $\mathbf{1}_A$ denotes the characteristic function of an open, half-open or closed interval A with end points $0 \leqslant \alpha < \beta \leqslant 1$. Then we have $\hat{f}(0) = (\beta - \alpha)$ and

$$\hat{\chi}(n) = \frac{1}{2\pi i n}\Big(\mathrm{e}(-n\beta) - \mathrm{e}(-n\alpha)\Big).$$

Moreover, we have $\hat{J}_{N+1}(0) = 1$ and $0 < \hat{J}_{N+1}(n) < 1$ for $n = 1, \ldots, N$. By Proposition A.21 and a simple continuity argument ($\chi(t)$ does not satisfy condition (A.27)) we obtain the following corollary:

Corollary A.22. *Let* $\chi(t)$ *be defined as above. Then there exist coefficients* $a_N(n) \in \mathbb{C}$ *with* $|a_N(n)| \leqslant 2$*, such that the trigonometric polynomial*

$$\chi_N^*(t) = (\beta - \alpha) + \frac{1}{2\pi i}\sum_{1\leqslant |n| \leqslant N} \frac{a_N(n)}{n}\,\mathrm{e}(nt)$$

verifies

$$|\chi(t) - \chi_N^*(t)| \leqslant \frac{1}{2N+2}\Big(\kappa_N\,(t-\beta) + \kappa_H\,(t-\alpha)\Big).$$

Bibliography

[AC85] Jean-Paul Allouche and Henri Cohen. Dirichlet series and curious infinite products. *Bull. London Math. Soc.*, 17(6):531–538, 1985.

[AL91] Jean-Paul Allouche and Pierre Liardet. Generalized Rudin-Shapiro sequences. *Acta Arith.*, 60(1):1–27, 1991.

[AL09] Isabelle Abou and Pierre Liardet. Flots chaînées. *Proceedings of the sixth Congress of Romanian Mathematicians, Bucharest, 2007, Scientific Contributions. Editors L. Beznea et al. Editura Academiei Române*, 1:401–432, 2009.

[AS93] Jean-Paul Allouche and Olivier Salon. Sous-suites polynomiales de certaines suites automatiques. *J. Théor. Nombres Bordeaux*, 5(1):111–121, 1993.

[AS03] Jean-Paul Allouche and Jeffrey Shallit. *Automatic sequences.* Cambridge University Press, Cambridge, 2003. Theory, applications, generalizations.

[AT00] Shigeki Akiyama and Jörg M. Thuswaldner. Topological properties of two-dimensional number systems. *J. Théor. Nombres Bordeaux*, 12(1):69–79, 2000.

[Bés72] Jean Bésineau. Indépendance statistique d'ensembles liés à la fonction "somme des chiffres". *Acta Arith.*, 20:401–416, 1972.

[BEW98] Bruce C. Berndt, Ronald J. Evans, and Kenneth S. Williams. *Gauss and Jacobi sums.* Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons Inc., New York, 1998. A Wiley-Interscience Publication.

[BI86] Enrico Bombieri and Henryk Iwaniec. On the order of $\zeta(\frac{1}{2}+it)$. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 13(3):449–472, 1986.

[Bil86] Patrick Billingsley. *Probability and measure.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, second edition, 1986.

[BK95] Nader L. Bassily and Imre Kátai. Distribution of the values of q-additive functions on polynomial sequences. *Acta Math. Hungar.*, 68(4):353–361, 1995.

[BK96] Nader L. Bassily and Imre Kátai. Distribution of consecutive digits in the q-ary expansions of some subsequences of integers. In *Proceedings of the XVI Seminar on Stability Problems for Stochastic Models, Part II (Eger, 1994)*, volume 78, pages 11–17, 1996.

[Des72] Jean-Marc Deshouillers. *Propriétés additive et arithmétique de suites à croissance polynômiale.* PhD thesis, Université de Paris VI, 1972.

[Des73a] Jean-Marc Deshouillers. Problème de Waring avec exposants non entiers. *Bull. Soc. Math. France*, 101:285–295, 1973.

[Des73b] Jean-Marc Deshouillers. Sur la répartition des nombres $[n^c]$ dans les progressions arithmétiques. *C. R. Acad. Sci. Paris Sér. A-B*, 277:A647–A650, 1973.

[DG] Michael Drmota and Peter J. Grabner. Analysis of digital functions and applications. Chapter 9 of *Combinatorics, Automata and Number Theory*, edited by Valérie Berthé and Michel Rigo, Cambridge University Press. To appear.

[DGL08] Michael Drmota, Peter J. Grabner, and Pierre Liardet. Block additive functions on the Gaussian integers. *Acta Arith.*, 135(4):299–332, 2008.

[DM10] Michael Drmota and Johannes F. Morgenbesser. Generalized Thue-Morse Sequences of Squares. *Israel Journal of Mathematics,* accepted, 2010.

[DMR] Michael Drmota, Christian Mauduit, and Joël Rivat. The sum of digits function of polynomial sequences. Submitted.

[DMR09] Michael Drmota, Christian Mauduit, and Joël Rivat. Primes with an average sum of digits. *Compos. Math.*, 145(2):271–292, 2009.

[DRS08] Michael Drmota, Joël Rivat, and Thomas Stoll. The sum of digits of primes in $\mathbb{Z}[i]$. *Monatsh. Math.*, 155(3-4):317–347, 2008.

[DT97] Michael Drmota and Robert F. Tichy. *Sequences, discrepancies and applications*, volume 1651 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 1997.

[DT06] Cécile Dartyge and Gérald Tenenbaum. Congruences de sommes de chiffres de valeurs polynomiales. *Bull. London Math. Soc.*, 38(1):61–69, 2006.

[FI89] Étienne Fouvry and Henryk Iwaniec. Exponential sums with monomials. *J. Number Theory*, 33(3):311–333, 1989.

[FM96] Étienne Fouvry and Christian Mauduit. Méthodes de crible et fonctions sommes des chiffres. *Acta Arith.*, 77(4):339–351, 1996.

[FM05] Étienne Fouvry and Christian Mauduit. Sur les entiers dont la somme des chiffres est moyenne. *J. Number Theory*, 114(1):135–152, 2005.

[Fog02] N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

[Gel68] Aleksandr O. Gelfond. Sur les nombres qui ont des propriétés additives et multiplicatives données. *Acta Arithmetica*, 13:259–265, 1968.

[Gil] William J. Gilbert. Gaussian Integers as Bases for Exotic Number Systems. Manuscript.

[Gil82] William J. Gilbert. Complex numbers with three radix expansions. *Canad. J. Math.*, 34(6):1335–1348, 1982.

[Gil87] William J. Gilbert. Complex bases and fractal similarity. *Ann. Sci. Math. Québec*, 11(1):65–77, 1987.

[GK91] Sidney W. Graham and Grigori Kolesnik. *Van der Corput's method of exponential sums*, volume 126 of *London Mathematical Society Lecture Note Series.* Cambridge University Press, Cambridge, 1991.

[GKP98] Peter J. Grabner, Peter Kirschenhofer, and Helmut Prodinger. The sum-of-digits function for complex bases. *J. London Math. Soc. (2)*, 57(1):20–40, 1998.

[GL66] Aleksandr O. Gelfond and Yuri V. Linnik. *Elementary methods in the analytic theory of numbers.* Translated from the Russian by D. E. Brown. Translation edited by I. N. Sneddon. International Series of Monographs in Pure and Applied Mathematics, Vol. 92. Pergamon Press, Oxford, 1966.

[GL99] Peter J. Grabner and Pierre Liardet. Harmonic properties of the sum-of-digits function for complex bases. *Acta Arith.*, 91(4):329–349, 1999.

[Gru07] Peter M. Gruber. *Convex and discrete geometry*, volume 336 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer, Berlin, 2007.

[GT00] Bernhard Gittenberger and Jörg M. Thuswaldner. Asymptotic normality of b-additive functions on polynomial sequences in the Gaussian number field. *J. Number Theory*, 84(2):317–341, 2000.

[Hla55] Edmund Hlawka. Zur formalen Theorie der Gleichverteilung in kompakten Gruppen. *Rend. Circ. Mat. Palermo (2)*, 4:33–47, 1955.

[HR95] Glyn Harman and Joël Rivat. Primes of the form $[p^c]$ and related questions. *Glasgow Math. J.*, 37(2):131–141, 1995.

[HST91] Edmund Hlawka, Johannes Schoissengeier, and Rudolf Taschner. *Geometric and analytic number theory.* Universitext. Springer-Verlag, Berlin, 1991. Translated from the 1986 German edition by Charles Thomas.

[IK04] Henryk Iwaniec and Emmanuel Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications.* American Mathematical Society, Providence, RI, 2004.

[Iwa78] Henryk Iwaniec. Almost-primes represented by quadratic polynomials. *Invent. Math.*, 47(2):171–188, 1978.

[Kát86] Imre Kátai. Distribution of digits of primes in q-ary canonical form. *Acta Math. Hungar.*, 47(3-4):341–359, 1986.

[Kim99] Dong-Hyun Kim. On the joint distribution of q-additive functions in residue classes. *J. Number Theory*, 74(2):307–336, 1999.

[KM68] Imre Kátai and József Mogyoródi. On the distribution of digits. *Publ. Math. Debrecen*, 15:57–68, 1968.

[KN74] Lauwerens Kuipers and Harald Niederreiter. *Uniform distribution of sequences.* Wiley-Interscience [John Wiley & Sons], New York, 1974. Pure and Applied Mathematics.

[Knu81] Donald E. Knuth. *The art of computer programming. Vol. 2.* Addison-Wesley Publishing Co., Reading, Mass., second edition, 1981. Seminumerical algorithms, Addison-Wesley Series in Computer Science and Information Processing.

[KS75] Imre Kátai and János Szabó. Canonical number-systems for complex integers. *Acta Sci. Math.*, 37:255–260, 1975.

[Lia90] Pierre Liardet. Automata and generalized Rudin-Shapiro sequences. *Arbeitsbericht Mathematisches Institut der Universität Salzburg*, 3-4:21–52, 1990.

[Mau01] Christian Mauduit. Multiplicative properties of the Thue-Morse sequence. *Period. Math. Hungar.*, 43(1-2):137–153, 2001.

[MMR] Bruno Martin, Christian Mauduit, and Joël Rivat. Sur les chiffres des nombres premiers. Preprint.

[Mor10a] Johannes F. Morgenbesser. The sum of digits of squares in $\mathbb{Z}[i]$. *J. Number Theory*, 130(7):1433–1469, 2010.

[Mor10b] Johannes F. Morgenbesser. The sum of digits of gaussian primes. Submitted, 2010.

[Mor10c] Johannes F. Morgenbesser. The sum of digits of $\lfloor n^c \rfloor$. Submitted, 2010.

[Mos07] Yossi Moshe. On the subword complexity of Thue-Morse polynomial extractions. *Theoret. Comput. Sci.*, 389(1-2):318–329, 2007.

[MR95] Christian Mauduit and Joël Rivat. Répartition des fonctions q-multiplicatives dans la suite $([n^c])_{n\in\mathbb{N}}$, $c > 1$. *Acta Arith.*, 71(2):171–179, 1995.

[MR05] Christian Mauduit and Joël Rivat. Propriétés q-multiplicatives de la suite $\lfloor n^c \rfloor$, $c > 1$. *Acta Arith.*, 118(2):187–203, 2005.

[MR09] Christian Mauduit and Joël Rivat. La somme des chiffres des carrés. *Acta Math.*, 203(1):107–148, 2009.

[MR10] Christian Mauduit and Joël Rivat. Sur un problème de Gelfond: la somme des chiffres des nombres premiers. *Annals of Mathematics*, 171(3):1591–1646, 2010.

[MRS02] Christian Mauduit, Joël Rivat, and András Sárközy. On the pseudo-random properties of n^c. *Illinois J. Math.*, 46(1):185–197, 2002.

[MS97] Christian Mauduit and András Sárközy. On the arithmetic structure of the integers whose sum of digits is fixed. *Acta Arith.*, 81(2):145–173, 1997.

[MTT01] Wolfgang Müller, Jörg M. Thuswaldner, and Robert F. Tichy. Fractal properties of number systems. *Period. Math. Hungar.*, 42(1-2):51–68, 2001.

[Pet03] Manfred Peter. The asymptotic distribution of elements in automatic sequences. *Theoret. Comput. Sci.*, 301(1-3):285–312, 2003.

[Rib96] Paulo Ribenboim. *The new book of prime number records.* Springer-Verlag, New York, 1996.

[Rie67] Georg J. Rieger. Über die natürlichen und primen Zahlen der Gestalt $[n^c]$ in arithmetischer Progression. *Arch. Math. (Basel)*, 18:35–44, 1967.

[RT05] Joël Rivat and Gérald Tenenbaum. Constantes d'Erdős-Turán. *Ramanujan J.*, 9(1-2):111–121, 2005.

[Ten08] Gérald Tenenbaum. *Introduction à la théorie analytique et probabiliste des nombres.* Éditions Belin, Paris, 2008.

[Vaa85] Jeffrey D. Vaaler. Some extremal functions in Fourier analysis. *Bull. Amer. Math. Soc. (N.S.)*, 12(2):183–216, 1985.

[Vin04] Ivan M. Vinogradov. *The method of trigonometrical sums in the theory of numbers.* Dover Publications Inc., Mineola, NY, 2004. Translated from the Russian, revised and annotated by K. F. Roth and Anne Davenport, Reprint of the 1954 translation.

[War74] Richard Warlimont. Über die Summe $\sum_{n\le x;\,[n^c]\equiv l(k)} 1$. *Arch. Math. (Basel)*, 25:151–153, 1974.