TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

D I S S E R T A T I O N

Correlations for numeration systems

Korrelationen für Zahlensysteme

Corrélations 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

Prof. Joël Rivat
Institut de
Mathématiques de Luminy
Université d'Aix-Marseille

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

Lukas SPIEGELHOFER

Faculté des sciences de Luminy
École Doctorale de Mathématiques et d'Informatique de Marseille
Institut de Mathématiques de Luminy

T H È S E

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ D'AIX-MARSEILLE

Spécialité: Mathématiques

par

Lukas SPIEGELHOFER

Titre:

Correlations for numeration systems

Corrélations pour les systèmes de numération

Korrelationen für Zahlensysteme

Les directeurs de thèse: Joël RIVAT et Michael DRMOTA
Les rapporteurs: Cécile DARTYGE et Gerhard LARCHER

Jury:

Cécile DARTYGE	Rapporteur
Michael DRMOTA	Directeur de thèse
Peter GRABNER	Examinateur
Joël RIVAT	Directeur de thèse

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 d'Aix-Marseille

ausgeführt.

Ce travail est présenté dans le cadre d'une

Convention de co-tutelle de thèse

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

Contents

Abstract in German, French and English

Deutsche Kurzfassung

Die vorliegende Dissertation behandelt Ziffernsummenfunktionen und die verwandten *q-additiven* und *q-multiplikativen* Funktionen. In den Arbeiten von C. Mauduit und J. Rivat zu dem sogenannten Gelfond-Problemen, die Ziffernsummenfunktionen betreffen, hat sich die diskrete Fouriertransformation als wertvolles Hilfsmittel erwiesen. Im ersten Kapitel wenden wir diese Technik im allgemeineren Rahmen von q-multiplikativen Funktionen an, was uns auf einen alternativen Beweis eines Resultates von J. Coquet, welches den Zusammenhang zwischen pseudozufälligen q-multiplikativen Funktionen und dem Fourier-Bohr-Spektrum betrifft, führt. Außerdem erhalten wir auf ähnliche Weise ein analoges Resultat für die Zeckendorf-Ziffernsumme.

In den nächsten beiden Kapiteln beschäftigen wir uns mit der Ziffernsumme von n und von $n+t$ und der Beziehung dieser Werte zueinander. Zunächst beweisen wir ein überraschendes Resultat über die Ziffernsumme von $n+t$ und $n+t^R$ (wobei t^R durch Umkehrung der Reihenfolge der Ziffern aus t hervorgeht). Danach erarbeiten wir eine partielle Antwort auf die folgende einfach zu formulierende Frage von T.W. Cusick: „Sei t eine natürliche Zahl. Ist es wahr, dass die binäre Ziffernsumme von $n+t$ für mehr als der Hälfte der natürlichen Zahlen mindestens so groß ist wie die binäre Ziffernsumme von n?“ Mit Hilfe von multivariaten erzeugenden Funktionen zeigen wir, dass die Antwort auf diese Frage für t in einer Teilmenge mit asymptotischer Dichte 1 positiv ist.

Das vierte Kapitel behandelt Teilfolgen natürlicher Zahlen von der Form $\lfloor n^c \rfloor$. Wir approximieren solche Folgen lokal durch Folgen einfacherer Gestalt, nämlich durch Beatty-Folgen $(\lfloor n\alpha + \beta \rfloor)_n$. Dieser Ansatz führt uns auf ein allgemeines Kriterium dafür, dass sich eine arithmetische Funktion auf $\lfloor n^c \rfloor$ so verhält, wie man es erwartet. Wir wenden dieses Theorem auf einige Probleme rund um Ziffernsummenfunktionen an.

Schließlich adaptieren wir die diskrete Fouriertransformation, um sie auf die Zeckendorf-Ziffernsumme anwenden zu können. Damit beweisen wir, dass sich diese Funktion und die gewöhnliche Ziffernsummenfunktion in Basis q (unter einer zusätzlichen Hypothese) unabhängig voneinander in Restklassen verteilen.

Résumé

Cette thèse porte sur la fonction bien connue de *somme des chiffres* et sur les notions associées de fonctions *q-additives* et *q-multiplicatives*. Dans les travaux de C. Mauduit et de J. Rivat sur les problèmes de Gelfond, qui ont pour objet la fonction somme de chiffres, l'utilisation de la transformée de Fourier discrète s'est avérée très efficace. Dans le premier chapitre, nous employons cette technique dans le contexte plus général des fonctions q-multiplicatives, ce qui nous permet de redémontrer un résultat de J. Coquet concernant la relation entre les fonctions q-multiplicatives pseudo-aléatoires et le spectre de Fourier-Bohr. De plus, nous obtenons un résultat analogue pour la fonction somme de chiffres associée au système de numération de Zeckendorf.

Les deux chapitres suivants sont consacrés à la relation entre la somme de chiffres de n et celle de $n+t$. Nous montrons d'abord une étonnante propriété de symétrie entre les chiffres et un résultat étroitement lié sur certaines suites 2-régulières. Ensuite, nous donnons une réponse partielle à la question de T. W. Cusick: «*Pour tout entier positif t, est-il vrai que la somme des chiffres de $n+t$ en base 2 est au moins aussi grand que la somme des chiffres de n en base 2 pour plus de la moitié de l'ensemble des entiers positifs n?*» En recourant aux fonctions génératrices de plusieurs variables, nous parvenons à montrer que la réponse est positive pour toutes les valeurs de t dans un ensemble de densité asymptotique égale à 1.

Dans le quatrième chapitre, nous nous intéressons aux sous-suites d'entiers de la forme $\lfloor n^c \rfloor$. Nous donnons des approximations locales de ces suites par des suites de Beatty, de forme plus simple $\lfloor n\alpha + \beta \rfloor$. Cela nous amène à établir un critère général permettant d'examiner si une fonction arithmétique évaluée aux entiers de la forme $\lfloor n^c \rfloor$ a le comportement attendu. Nous appliquons notre théorème à certains problèmes en lien avec les fonctions sommes de chiffres.

Enfin, nous adaptons la technique de la transformée de Fourier discrète à la fonction somme des chiffres dans le système de numération de Zeckendorf afin de montrer que cette fonction ainsi que la fonction somme des chiffres en base q sont indépendemment distribuées dans les classes d'équivalence (sous des hypothèses assez faibles).

Abstract

This thesis is concerned with the well-known sum-of-digits function and the related notions of *q-additive* and *q-multiplicative* functions. In the work of C. Mauduit and J. Rivat on the so-called Gelfond problems, which deal with the sum-of-digits function, the discrete Fourier transform has proven to be a valuable tool. In the first chapter we apply this technique to the more general context of q-multiplicative functions, thereby reproving a result of J. Coquet concerning the relation of pseudorandom q-multiplicative functions to the Fourier-Bohr spectrum. Moreover, in this vein we establish an analogous result for the Zeckendorf sum-of-digits function.

In the next two chapters we are concerned with the relation of the sum of digits of n and $n+t$ to each other. We first prove a surprising digit reflection property and a related result on certain 2-regular sequences. After that, we establish a partial answer to the following simple-to-state question by T. W. Cusick: *for each positive integer t, is it true that the binary sum of digits $n+t$ is at least as large as the binary sum of digits of n for more than half of all positive integers n?* Using multivariate generating functions, we show that the answer to this question is positive for t in a set of asymptotic density 1.

The fourth chapter deals with subsequences of the integers of the form $\lfloor n^c \rfloor$. We locally approximate this kind of sequence by sequences of the simpler form $\lfloor n\alpha + \beta \rfloor$ (so-called Beatty sequences). This leads us to a general criterion for an arithmetic function evaluated on integers of the form $\lfloor n^c \rfloor$ to behave as expected. We apply this theorem to certain problems related to sum-of-digits functions.

Finally, we adapt the discrete Fourier transform technique to the Zeckendorf sum-of-digits function in order to prove that this function and the sum-of-digits function in base q are (under some mild assumption) distributed independently in residue classes.

Chapter 0

Introduction in German, French and English

0.1 Deutsche Einleitung

Zu Beginn dieser Arbeit definieren wir die wichtigsten Begriffe, die in den Kapiteln 1 bis 5 benötigt werden. Der zentrale Begriff ist die *Zifferndarstellung* einer ganzen Zahl n, auch *Zahlensystem* genannt. In einem sehr allgemeinen Kontext ist ein solches System nichts anderes als eine injektive Abbildung von den natürlichen Zahlen in eine Menge von Folgen von *Ziffern*. Das wohl bekannteste Zahlensystem ist das Dezimalsystem, das von einem nicht unwesentlichen Teil der Erdbevölkerung in vielen Lebensbereichen verwendet wird. Etwas weniger bekannt, jedoch zumindest innerhalb der Mathematik nicht weniger wichtig, ist die Verallgemeinerung dieses Systems auf beliebige ganzzahlige Basen $q \geq 2$: Jede nichtnegative ganze Zahl n lässt sich in eindeutiger Weise als

$$n = \sum_{i \geq 0} \varepsilon_i q^i$$

schreiben, wobei $\varepsilon_i \in \{0, \ldots, q-1\}$ und $\varepsilon_i \neq 0$ nur endlich oft. Vor Allem das Dualsystem ($q = 2$) hat im vergangenen Jahrhundert einen bemerkenswerten Aufstieg erlebt, was auf die Erfindung digitaler Rechenmaschinen zurückzuführen ist. Die Darstellung einer ganzen Zahl in Basis q wird uns in jedem der fünf Kapitel dieser Arbeit begegnen. Ein weiteres Zahlensystem wird uns an mehreren Stellen begegnen, nämlich die *Zeckendorf-Entwicklung* einer ganzen Zahl n. Dieses System ist auf Zeckendorf's Theorem basiert, welches besagt, dass jede positive ganze Zahl n in eindeutiger Weise als Summe von Fibonaccizahlen geschrieben werden kann, wobei allerdings von je zwei benachbarte Fibonaccizahlen nicht beide vorkommen dürfen. (Es ist einfach zu sehen, dass diese Bedingung notwendig ist, denn falls sowohl F_k als auch F_{k+1} in einer Darstellung auftreten, wobei k maximal ist, so kann man statt dessen F_{k+2} nehmen, was eine weitere Darstellung liefert.) In anderen Worten kann man jede nichtnegative ganze Zahl n in eindeutiger Weise als

$$n = \sum_{k \geq 0} \varepsilon_k F_k$$

schreiben, wobei $\varepsilon_k \in \{0,1\}$ und $\varepsilon_k \neq 0$ nur endlich oft, und falls $\varepsilon_k = 1$, dann $\varepsilon_{k+1} = 0$. Dieses Zahlensystem ist ein Spezialfall des Ostrowski-Zahlensystems, das auf der Kettenbruchdarstellung einer reelen Zahl basiert ist. Die Zeckendorf-Entwicklung erhält man daraus, in dem man in Bezug auf den goldenen Schnitt φ entwickelt. Wir werden diesen allgemeinen Fall nicht betrachten, sondern geben uns mit der Hoffnung darauf zufrieden, dass man manche unserer Beweise auf den allgemeinen Fall übertragen kann. Basierend auf den eben definierten Zahlensystemen können wir Ziffernsummenfunktionen definieren. Gegeben eine ganze Zahl $q \geq 2$. Dann sei s_q diejenige Funktion, die die Ziffern von n in der q-adischen Darstellung summiert. Außerdem sei Z diejenige Funktion, die die Anzahl der Summanden in der Zeckendorf-Entwicklung zurückgibt. In den vergangenen Jahren wurde der Ziffernsummenfunktion s_q einiges an Aufmerksamkeit geschenkt und große Fortschritte an den so genannten Gelfond-Problemen gemacht. Diese Probleme werden am Ende des Artikels [29] von Gelfond vorgeschlagen und besagen das Folgende:

1. Studiere die gemeinsame Verteilung von Ziffernsummenfunktionen in verschiedenen Basen in Restklassen.

2. Finde die Anzahl der Primzahlen $p \leq x$ mit $s_q(p) \equiv \ell \bmod m$.

3. Für ein Polynom P, das $P(n) \in \mathbb{N}$ für $n \in \mathbb{N}$ erfüllt, studiere die Verteilung von $s_q(P(n))$ in Restklassen.

Das erste dieser Probleme wurde von Kim [33] vollständig gelöst. Das zweite Problem und der Spezialfall $P(n) = n^2$ des dritten wurde von Mauduit und Rivat [41, 42] gelöst.

Wir betrachten auch Verallgemeinerungen der Ziffernsummenfunktion s_q und der zugehörigen Funktion $n \mapsto \mathrm{e}(\vartheta s_q(n))$ (wobei hier und im Rest dieser Arbeit $\mathrm{e}(x) = \exp(2\pi i x)$):

Eine arithmetische Funktion f heit q-additiv, falls es Funktionen $f_k : \{0, \dots, q-1\} \to \mathbb{C}$ gibt, die $f_k(0) = 0$ und

$$f\left(\sum_{k\geq 0} a_k q^k\right) = \sum_{k\geq 0} f_k(a_k)$$

erfüllen, wobei $a_k \in \{0, \dots, q-1\}$ und $a_k \neq 0$ nur endlich oft. Diese Bedingung ist dazu äquivalent, dass $f(q^k n + b) = f(q^k n) + f(b)$ für alle nichtnegativen ganzen Zahlen k, n, die $0 \leq b < q^k$ erfüllen. Das ist auch die Definition, die in Kapitel 1 gegeben wird. Analog heißt g *q-multiplikativ*, falls es Funktionen $g_k : \{0, \dots, q-1\} \to \mathbb{C}$ mit $g_k(0) = 1$ gibt, die

$$g\left(\sum_{k\geq 0} a_k q^k\right) = \prod_{k\geq 0} g_k(a_k)$$

erfüllen, wobei $a_k \in \{0, \dots, q-1\}$ und $a_k \neq 0$ nur endlich oft. Für den Fall dass $q = 2$ wurde der Begriff „q-additiv“in Bellman and Shapiro [5] definiert (wo diese Funktionen *dyadisch additiv* heißen). Seitdem haben viele Autoren diese beiden Typen von Funktionen studiert. Wir verweisen den Leser auf [40], worin zahlreiche Verweise auf die Literatur angegeben werden. Wir bemerken noch, dass die Ziffernsummenfunktion s_q aus der allgemeinen Definition einer q-additiven Funktion hervorgeht, indem man $f_k(b) = b$ setzt.

Kapitel 1 erfüllt einen zweifachen Zweck. Erstens ist es ein einführendes Kapitel, das lange zuvor bewiesene Resultate über q-additive und q-multiplikative Funktionen Revue passieren lässt (und auch eine Charakterisierung des Verhaltens des Mittelwertes von q-additiven Funktionen liefert, was zwar in dieser Form neu, aber nicht tiefschürfend ist). Die wichtigste Referenz für diesen Teil ist Delange [16]. Der zweite und wichtigere Zweck dieses Kapitels ist es zu zeigen, dass die Verwendung der diskreten Fourier-Transformation ein nützliches Werkzeug ist, um q-multiplikative Funktionen zu studieren. Es ist eine der grundlegende Methoden in den beiden oben zitierten Artikeln von Mauduit und Rivat, die diskrete Fouriertransformation für die Behandlung von Ziffernsummenproblemen zu verwenden. Wir zeigen, dass diese Methode auch in dem allgemeineren Rahmen sinnvoll anwendbar ist. Für $\lambda \geq 0$ und $h \in \mathbb{Z}$ haben die diskreten Fourierkoeffizienten der q^λ-periodischen Funktion $u \mapsto \mathrm{e}\left(\vartheta s_q(u \bmod q^\lambda)\right)$ die Form

$$F_\lambda(h,\vartheta) = \frac{1}{q^\lambda} \sum_{u<q^\lambda} \mathrm{e}\left(\vartheta s_q(u) - huq^{-\lambda}\right).$$

Indem wir analoge Ausdrücke für q-multiplikative Funktionen verwenden, beweisen wir auf einfachere Weise ein Theorem von Coquet [10], welches eine Verbindung zwischen dem *Fourier-Bohr Spektrum* einer q-multiplikativen Funktion g und einem bestimmten Begriff von *Pseudozufälligkeit* herstellt. Dabei ist das Fourier-Bohr Spektrum einer arithmetischen Funktion g die Menge der $\beta \in [0,1)$, die

$$\limsup_{N\to\infty} \frac{1}{N} \left| \sum_{n<N} g(n)\,\mathrm{e}\left(\beta n\right) \right| > 0$$

erfüllen. Eine arithmetische Funktion g heißt pseudozufällig (in Sinne von Bertrandias), wenn die (Auto)korrelation

$$\gamma_t = \lim_{N\to\infty} \frac{1}{N} \sum_{n<N} g(n+t)\overline{g(n)}$$

für alle $t \geq 0$ existiert und das quadratische Mittel der γ_t gleich 0 ist. Der schwierigere Teil des oben genannten Theorems von Coquet ist der folgende Satz.

Theorem (Coquet)**.** *Sei f eine q-additive Funktion und $g(n) = \mathrm{e}\left(f(n)\right)$. Wenn das Fourier-Bohr Spektrum von g leer ist, dann ist g pseudozufällig.*

Unser Beweis dieses Theorems besteht in einer Anwendung von van der Corput's Ungleichung, gefolgt von einer diskreten Fouriertransformation. Es ist das die Kombination, die sich auch in den Artikeln [41, 42] als erfolgreich erwiesen hat. Außerdem beweisen wir mit denselben Methoden ein verwandtes Resultat für die arithmetische Funktion $n \mapsto \mathrm{e}\left(\vartheta Z(n)\right)$.

Theorem. *Sei $\vartheta \in \mathbb{R} \setminus \mathbb{Z}$. Dann ist die Funktion $n \mapsto \mathrm{e}\left(\vartheta Z(n)\right)$ pseudozufällig.*

In **Kapitel 2** betrachten wir die Autokorrelation $\gamma_t(\vartheta)$ der q-multiplikativen Funktion $n \mapsto \mathrm{e}\left(\vartheta s_q(n)\right)$ etwas näher. Diese Folge ist ein Beispiel einer sogenannten q-regulären Folge

(siehe beispielsweise Drmota und Grabner [21]). Es gilt die folgende Identität: Ist $(a_\nu \cdots a_0)_q$ die Darstellung von t in Basis q, dann

$$\gamma_t(\vartheta) = (1\,,0)\,A(a_0)\cdots A(a_\nu)\begin{pmatrix}1\\u\end{pmatrix}$$

für gewisse $u \in \mathbb{C}$ und 2×2-matrizen $A(0),\ldots,A(q-1)$. Mithilfe dieser Formel beweisen wir, dass die Korrelationen für die Ziffernsummenfunktion eine Symmetrie bezüglich *Spiegelung der Ziffern* besitzen: Für nichtnegative ganze Zahlen t sei t^R diejenige Zahl, die durch Schreiben der Ziffern von t in umgekehrter Reihenfolge entsteht. Mit dieser Notation beweisen wir in [47] das folgende Theorem.

Theorem (Morgenbesser und Spiegelhofer)**.** *Für $q \geq 2$, $\vartheta \in \mathbb{R}$ und $t \geq 0$ sei*

$$\gamma_t(\vartheta) = \lim_{N\to\infty}\frac{1}{N}\sum_{n<N} e(\vartheta(s_q(n+t)-s_q(n)).$$

Dann gilt

$$\gamma_t(\vartheta) = \gamma_{t^R}(\vartheta),$$

wobei die Ziffern-Umkehrung t^R in Bezug auf die Basis q zu nehmen ist.

Dieses Resultat ist unerwartet, denn die Ziffernsumme von $n+t$ scheint nichts mit der Ziffernsumme von $n+t^R$ zu tun zu haben. Allerdings ist der Beweis nicht lang, sobald man eine adequate Induktionshypothese gefunden hat.[1] Diese Induktionshypothese ist der Bestandteil des Beweises, der vom Himmel fällt, während der Rest aus linearer Algebra besteht.

Im Fall $q = 2$ zeigen wir ein allgemeineres Theorem.

Theorem. *Es seien α und β komplexe Zahlen. Weiters sei $(z_n)_{n\geq 0}$ eine Folge mit der Eigenschaft*

$$z_{2t} = z_t \quad \textit{and} \quad z_{2t+1} = \alpha z_t + \beta z_{t+1}$$

für alle $t \geq 1$. Dann gilt $z_{t^R} = z_t$, wobei die Ziffern in Bezug auf die Basis $q = 2$ gespiegelt werden.

Wieder ist ein Induktionsbeweis im Spiel und der einzig schwierige Teil ist es, die Induktionshypothese zu finden.

In **Kapitel 3** setzen wir unser Studium der Beziehung der Ziffernsummen von n und $n+t$ zueinander fort. Die folgende Frage [14] steht im Raum: Gilt für jedes $t \geq 0$, dass die Menge der n mit $s_2(n+t) \geq s_2(n)$ asymptotische Dichte $c_t > 1/2$ hat? Auf den ersten Blick mag es aussehen, als wäre diese Eigenschaft einfach zu beweisen. Allerdings lässt eine eingehendere Betrachtung das Gegenteil vermuten – diese Frage kann beispielsweise in ein Problem über die Teilbarkeitseigenschaften von Binomialkoeffizienten übersetzt werden, was sich als ein sehr schwierig zu behandelndes Thema herausgestellt hat.

Als erstes beweisen wir, dass „für die Hälfte“ der $t \in \mathbb{N}$ die nichttriviale untere Schranke $c_t > 15/32$ gilt. Danach betrachten wir c_t für t von der speziellen Form $t = ((10)^j)_2$, wobei wir zeigen, dass diese Werte größer als $1/2$ sind, sobald j groß genug ist. Das Hauptresultat dieses Kapitels ist allerdings eine Aussage über „fast alle“ t.

[1] Der Autor hat den Fall $q = 2$ bewiesen, während der kritische Schritt im Beweis der allgemeinen Aussage, Proposition 2.10, von J. Morgenbesser ausfindig gemacht wurde.

Theorem. *Für $T \to \infty$ gilt*

$$T - |\{t \le T : c_t > 1/2\}| = O\left(\frac{T}{\sqrt{\log T}}\right),$$

das bedeutet, $c_t > 1/2$ gilt für t in einer Teilmenge $\mathbb{N}$ mit asymptotischer Dichte 1.

Der Beweis basiert auf der Untersuchung der Konzentration der Werte c_t um den Erwartungswert, das heißt, auf der Methode des zweiten Moments. Wir wenden diese Methode auf die Folge $(X_\lambda)_\lambda$ von Zufallsvariablen $X_\lambda : t \mapsto c_t$ an, wobei t im endlichen Intervall $[2^\lambda, 2^{\lambda+1})$ enthalten ist. Der Erwartungswert ist schnell bestimmt und liegt über $1/2$. Währenddessen zeigt sich, dass die Folge der zweiten Momente um einiges schwerer zu fassen ist – wir charakterisieren diese Folge als die *Diagonale* einer rationalen Funktion in drei Variablen. Wir extrahieren den Koeffizienten $[x^n y^n z^n]$ mithilfe der mehrdimensionalen Cauchyschen Integralformel und der Sattelpunktsmethode.

Kapitel 4 ist eine annähernd unveränderte Version des Artikels "Piatetski-Shapiro sequences via Beatty sequences", der für die Publikation in den Acta Arithmetica akzeptiert wurde. (Dabei erscheinen einige der Lemmata der vorhergehenden Kapitel ein zweites Mal, was dieses Kapitel unabhängig macht.)

Wir approximieren Folgen der Form $(\lfloor n^c \rfloor)_n$ (Piatetski-Shapiro-Folgen) lokal durch Beatty-Folgen $(\lfloor n\alpha + \beta \rfloor)_n$, was im Wesentlichen Taylorsche Approximation ist, und beweisen so das folgende Theorem.

Theorem. *Sei f eine zweimal stetig differenzierbare reellwertige Funktion auf $\mathbb{R}^+$ mit f, f', $f'' > 0$. Seien $c_1 \ge 1/2$ und $c_2 > 0$ so, dass für $0 < x \le y \le 2x$ die Formel $c_1 f''(x) \le f''(y) \le c_2 f''(x)$ gilt. Weiters sei $A_0 \ge 2$ dergestalt dass $f'(A_0) \ge 1$. Es existiert eine Konstante $C = C(f)$ sodass für alle komplexwertigen durch 1 beschränkten arithmetischen Funktionen φ, für alle ganzen Zahlen $A \ge A_0$ und für alle $z > 0$ die Abschätzung*

$$\frac{1}{A}\left|\sum_{A<n\le 2A} \varphi\left(\lfloor f(n) \rfloor\right) - \sum_{f(A)<m\le f(2A)} \varphi(m)\left(f^{-1}\right)'(m)\right| \le C\left(\frac{f''(A)}{f'(A)^2} z^2 + f'(A)(\log A)^3 J(A,z)\right) \quad (1)$$

gilt, wobei

$$J(A,z) = \int_0^1 \sup_{f(A)<x\le f(2A)} \frac{1}{z}\left|\sum_{x<m\le x+z} \varphi(m)\,\mathrm{e}\left(m\vartheta\right)\right| \mathrm{d}\vartheta. \quad (2)$$

Als Korollar liefert dieses Theorem eine hinreichende Bedingung dafür dass eine arithmetische Funktion φ ausgewertet an $\lfloor n^c \rfloor$ sich so verhält „wie man es erwartet". Mithilfe dieses Resultates studieren wir das Verhalten der Thue-Morse-Folge

$$(0,1,1,0,1,0,0,1,1,0,0,1,0,1,1,0,\ldots)$$

auf $\lfloor n^c \rfloor$, wobei wir den Bereich der zulässigen Exponenten c von $(1;1,4)$ (siehe [40]) auf $(1;1,42]$ erweitern:

Theorem. *Für* $1 < c \leq 1,42$ *existieren* $\eta > \max\{0, (7-5c)/9\}$ *und eine Konstante* C*, sodass für alle* $N \geq 2$

$$\frac{1}{N}\left|\sum_{1\leq n\leq N}(-1)^{s_2(\lfloor n^c\rfloor)}\right| \leq CN^{-\eta}.$$

Wir wenden das Hauptresultat auf zwei weitere Situationen an. Die zweite Anwendung hat die gemeinsame Verteilung in Restklassen von Ziffernsummenfunktionen zu verschiedenen Basen auf $\lfloor n^c\rfloor$ zum Thema. Wir zeigen ein Unabhängigkeitsresultat, indem wir erneut die Fourierkoeffizienten $F_\lambda(h,\vartheta)$ ins Spiel bringen. Wir bemerken auch, dass diese Methode einen alternativen Beweis des ersten Gelfond-Problems liefert, welches nichts Anderes ist als der Fall $c = 1$. Die dritte Anwendung betrifft die Verteilung der Zeckendorf-Ziffernsumme von $\lfloor n^c\rfloor$ in Restklassen. In beiden Fällen erhalten wir einen nichttrivialen Exponenten c.

In **Kapitel 5** studieren wir die gemeinsame Verteilung der Funktionen s_q and Z in Restklassen. Wie oben bemerkt, wurde die gemeinsame Verteilung von s_{q_1} und s_{q_2} bereits von Kim [33] behandelt, der ein quantitatives Unabhängigkeitsresultat bewiesen hat. Auch die gemeinsame Verteilung von s_q und Z ist behandelt worden (siehe Coquet, Rhin und Toffin [13]), allerdings ohne einen Fehlerterm zu liefern. Der Zweck dieses Kapitels ist es, das folgende Theorem zu beweisen.

Theorem. *Sei* $q \geq 2$ *eine ganze Zahl und* ϑ, β *reele Zahlen mit* $\beta \notin \mathbb{Z}$*. Dann gilt*

$$\sum_{n<N}\mathrm{e}\left(\vartheta s_q(n) + \beta Z(n)\right) = O\left(N^{1-\eta}\right)$$

für ein $\eta > 0$*.*

Um das zu beweisen, charakterisieren wir die natürlichen Zahlen, deren Zeckendorf-Entwicklung mit einer gegebenen Folge von Ziffern in $\{0,1\}$ beginnt. Das entsprechende Problem für die Entwicklung in Basis q ist einfach: Die Entwicklung von n in Basis q beginnt mit einer bestimmten Folge $(a_0,\ldots,a_{k-1})$ von Ziffern genau dann, wenn n in einer entsprechenden Restklasse modulo q^k liegt. Dank dieser Tatsache kann man, zum Beispiel im Zusammenhang mit den Gelfond-Problemen, die diskrete Fouriertransformation sinvoll einsetzen. Um eine analoge Aussage für die Zeckendorf-Entwicklung zu erlangen, nützen wir die Tatsache aus, dass diejenigen $n \in \mathbb{N}$, deren erste Ziffern in dieser Entwicklung fixiert sind, durch die Bedingung $n\varphi \in I + \mathbb{Z}$ charakterisiert sind, wobei $\varphi = (\sqrt{5}+1)/2$ und I ein Intervall ist, das zu den ersten Ziffern gehört. Indem wir diese Eigenschaft ausnützen, erhalten wir ein Resultat, das zur inversen diskreten Fouriertransformation analog ist (siehe Proposition 5.4). Mithilfe dieses Resultates beweisen wir das Theorem.

0.2 Introduction en français

Nous commençons ce travail en introduisant les notions fondamentales qui sont utilisées dans les chapitres 1 à 5. Le concept central de cette thèse est la représentation d'un nombre entier n dans une base donnée q, également connu comme système de numération. Dans un cadre très général, un tel système est simplement une fonction injective des nombres entiers positifs vers un ensemble de suites de chiffres.

Le système de numération le plus connu est le système décimal ($q = 10$), utilisé dans la vie quotidienne par une partie importante de la population du monde. Un peu moins connue, mais tout de même importante, est la généralisation de ce système à des bases q arbitraires, où $q \geq 2$ est un nombre entier: tout entier n non négatif peut s'écrire de manière unique sous la forme

$$n = \sum_{i \geq 0} \varepsilon_i q^i,$$

où $\varepsilon_i \in \{0, \ldots, q-1\}$ et $\varepsilon_i = 0$ pour tout i sauf un nombre fini d'exceptions. En particulier le cas $q = 2$ a massivement pris de l'importance pour l'humanité dans le siècle dernier, en raison de l'invention des machines de calcul numérique. La notion de la représentation d'un entier dans une base q sera présente dans chaque chapitre de cette thèse. Il y a un autre système qui va apparaître à plusieurs endroits, c'est la *représentation de Zeckendorf* d'un entier n. Elle est basée sur le théorème de Zeckendorf, indiquant que chaque entier positif n peut s'écrire de manière unique comme une somme de nombres de Fibonacci, donnés par $1, 2, 3, 5, 8, 13, \ldots$, où il est interdit de prendre deux nombres de Fibonacci adjacents. (Il est facile de voir que cette dernière condition est nécessaire, car si F_k et F_{k+1} apparaîssent tous les deux dans une représentation de n comme une somme de nombres de Fibonacci, où k est maximal, en prenant F_{k+2} on obtient une autre représentation.) En d'autres termes, chaque entier naturel n peut s'écrire de manière unique comme

$$n = \sum_{k \geq 0} \varepsilon_k F_k,$$

où $\varepsilon_k \in \{0, 1\}$ et $\varepsilon_k = 0$ pour tout k sauf un nombre fini, et si $\varepsilon_k = 1$, alors $\varepsilon_{k+1} = 0$. Ce système est un cas particulier du système de numération dit d'Ostrowski, qui est basé sur le développement en fraction continue d'un nombre réel. Le cas de la représentation Zeckendorf est obtenu en prenant la numération Ostrowski par rapport au nombre d'or φ. Toutefois, nous ne serons pas concernés par le cas général. Nous mentionnons que certaines preuves de nos résultats concernant la représentation Zeckendorf pourraient être adaptées au cas plus général de la numération Ostrowski, donnant des résultats analogues. Sur la base des représentations digitales ci-dessus, nous pouvons définir des fonctions sommes de chiffres: pour tout entier $q \geq 2$ soit s_q la fonction qui additionne les chiffres de n dans la représentation en base q. En outre, soit Z la fonction donnant le nombre de termes de la somme dans la représentation Zeckendorf de n. Au cours des dernières années, la fonction somme des chiffres s_q a suscité un certain intérêt et de grands progrès ont été réalisés sur les problèmes connus sous le nom de Gelfond. Ces problèmes sont proposés à la fin de l'article [29] par Gelfond et peuvent se formuler ainsi:

1. Étudier la distribution conjointe dans les classes de résidus de la fonction somme de chiffres dans des bases différentes.

2. Trouver le nombre de nombres premiers $p \leq x$ tels que $s_q(p) \equiv \ell \bmod m$.

3. Pour tout polynôme P tel que $P(n) \in \mathbb{N}$ pour $n \in \mathbb{N}$, étudier la distribution de $s_q(P(n))$ dans les classes de congruence.

Le premier de ces problèmes a été complètement résolu par Kim [33]. Le deuxième problème et le cas particulier $P(n) = n^2$ du troisième problème ont été résolus par Mauduit et Rivat [41, 42]. Nous généralisons la fonction somme des chiffres s_q dans la base q et la fonction correspondante $n \mapsto \mathrm{e}(\vartheta s_q(n))$ (où $\mathrm{e}(x) = \exp(2\pi i x)$) en définissant les termes «fonction q-additive» et «fonction q-multiplicative» comme suit. Une fonction arithmétique f est appelée *q-additive* si il y a des fonctions $f_k : \{0, \dots, q-1\} \to \mathbb{C}$ telles que $f_k(0) = 0$ et

$$f\left(\sum_{k\geq 0} a_k q^k\right) = \sum_{k\geq 0} f_k(a_k),$$

où $a_k \in \{0, \dots, q-1\}$ et a_k est égal à zéro pour tout k sauf un nombre fini

Cette condition est équivalente à la condition que $f(q^k n + b) = f(q^k n) + f(b)$ quand k, n sont des entiers positifs tels que $0 \leq b < q^k$, ce qui est la définition donnée au chapitre 1.

De manière analogue, g est appelée *q-multiplicative* si il y a des fonctions $g_k : \{0, \dots, q-1\} \to \mathbb{C}$ telles que $g_k(0) = 1$ et

$$g\left(\sum_{k\geq 0} a_k q^k\right) = \prod_{k\geq 0} g_k(a_k),$$

où $a_k \in \{0, \dots, q-1\}$ et a_k est égal à zéro pour presque tout k. Pour le cas où $q = 2$, le terme fonction q-additive a été défini dans Bellman et Shapiro [5] (où ces fonctions sont appelées *dyadically additive*). De nombreux auteurs ont étudié depuis les fonctions q-additives et q-multiplicatives. Le lecteur trouvera de nombreuses références à la littérature dans [40]. La fonction somme des chiffres découle de la définition générale en fixant $f_k(b) = b$. Le but du **chapitre 1** est double. Tout d'abord, il s'agit d'un chapitre d'introduction. Nous passons en revue quelques résultats déjà prouvés sur les fonctions q-additives et q-multiplicatives et donnons aussi une caractérisation du comportement de la valeur moyenne des fonctions q-additives. La référence principale dans la littérature pour cette partie est Delange [16].

Plus important encore, nous voulons montrer que la transformée de Fourier discrète est un outil précieux dans l'étude des fonctions q-multiplicatives. L'utilisation de la transformée de Fourier dans le cadre de la fonction somme de chiffres est l'une des techniques fondamentales dans les deux articles de Mauduit et Rivat cités ci-dessus, et on montre que cette technique est également utile dans le contexte plus général.

Pour $\lambda \geq 0$ et $h \in \mathbb{Z}$ donnés, les coefficients de la transformation de Fourier discrète pour la fonction $u \mapsto \mathrm{e}\left(\vartheta s_q(u \bmod q^\lambda)\right)$, périodique de la période q^λ, ont la forme

$$F_\lambda(h, \vartheta) = \frac{1}{q^\lambda} \sum_{u<q^\lambda} \mathrm{e}\left(\vartheta s_q(u) - huq^{-\lambda}\right).$$

En utilisant des coefficients de Fourier pour les fonctions q-multiplicatives, nous redémontrons de façon plus simple un théorème de Coquet [10] concernant la relation entre le *spectre de Fourier-Bohr* d'une fonction q-multiplicative g et une certaine notion de fonction pseudo-aléatoire.

Le spectre de Fourier-Bohr d'une fonction arithmétique g est l'ensemble des $\beta \in [0, 1)$ tels

que

$$\limsup_{N\to\infty} \frac{1}{N} \left| \sum_{n<N} g(n)\, \mathrm{e}\,(\beta n) \right| > 0.$$

Une fonction arithmétique g est appelé pseudo-aléatoire (au sens de Bertrandias) si la corrélation (l'autocorrélation)

$$\gamma_t = \lim_{N\to\infty} \frac{1}{N} \sum_{n<N} g(n+t)\overline{g(n)}$$

existe pour $t \geq 0$ et, de plus, la moyenne quadratique de γ_t est égal à 0. La partie critique du théorème par Coquet précité est le théorème suivant.

Théorème (Coquet)**.** *Soit f une fonction q-additive et $g(n) = \mathrm{e}\,(f(n))$. Si le spectre de Fourier-Bohr de g est vide, g est pseudo-aléatoire.*

Notre preuve de ce théorème est une application de l'inégalité de van der Corput, suivie d'une transformée de Fourier discrète. C'est la combinaison qui a fait ses preuves dans [41, 42]. Nous prouvons aussi un résultat lié pour la fonction arithmétique $n \mapsto \mathrm{e}\,(\vartheta Z(n))$.

Théorème. *Soit $\vartheta \in \mathbb{R} \setminus \mathbb{Z}$. La fonction $n \mapsto \mathrm{e}\,(\vartheta Z(n))$ est pseudo-aléatoire.*

Dans le **chapitre 2** nous étudions l'auto-corrélation $\gamma_t(\vartheta)$ de la fonction q-multiplicative $n \mapsto \mathrm{e}\,(\vartheta s_q(n))$. Cette suite est un exemple de suite dite q-régulière (voir Drmota et Grabner [21]) et nous avons la représentation suivante: si $(a_\nu \cdots a_0)_q$ est la représentation de t en base q, alors

$$\gamma_t(\vartheta) = (1\,,0)\, A(a_0) \cdots A(a_\nu) \begin{pmatrix} 1 \\ u \end{pmatrix}$$

pour certains $u \in \mathbb{C}$ et certaines matrices 2×2 $A(0)$ to $A(q-1)$.

En utilisant cette formule, nous montrons que les corrélations pour la fonction somme des chiffres satisfont une propriété de symétrie par rapport à l'*inversion des chiffres*: pour tout nombre entier non négatif t, soit t^R l'entier obtenu en écrivant les chiffres dans la base q dans l'ordre inverse. Dans [47] nous prouvons le théorème suivant:

Théorème (Morgenbesser et Spiegelhofer)**.** *Pour $q \geq 2$, $\vartheta \in \mathbb{R}$ et $t \geq 0$ soit*

$$\gamma_t(\vartheta) = \lim_{N\to\infty} \frac{1}{N} \sum_{n<N} e(\vartheta(s_q(n+t) - s_q(n)).$$

Alors, nous avons

$$\gamma_t(\vartheta) = \gamma_{t^R}(\vartheta),$$

où la réflexion t^R doit être prise par rapport à la base q.

Ce résultat est inattendu dans la mesure où la somme des chiffres de $n+t$ semble être sans rapport avec la somme des chiffres de $n+t^R$. La preuve de cette assertion n'est pas longue sous une hypothèse de récurrence adéquate.[2] Trouver cette hypothèse est la partie non évidente de la preuve, alors que le reste est de l'algèbre linéaire. Dans le cas particulier où $q = 2$, nous établissons un théorème plus général.

[2]L'auteur a trouvé une preuve pour le cas où $q = 2$, alors que l'étape critique dans le cas général, ce qui est la proposition 2.10, est due à J. Morgenbesser.

Théorème. *Soient α et β des nombres complexes et $(z_n)_{n\geq 0}$ une suite satisfaisant la récurrence*

$$z_{2t} = z_t \quad et \quad z_{2t+1} = \alpha z_t + \beta z_{t+1}$$

pour tout $t \geq 1$. Alors $z_{t^R} = z_t$, où l'inversion de chiffres est par rapport à la base 2.

Encore une fois la preuve se fait par récurrence et la partie la plus délicate est de trouver l'hypothèse de récurrence.

Dans le **chapitre 3** nous continuons à étudier le rapport de la somme des chiffres de n et $n+t$. Il est admis [14] que pour chaque t, la densité asymptotique c_t de l'ensemble des n tels que $s_2(n+t) \geq s_2(n)$ satisfait $c_t > 1/2$. A première vue, cela peut sembler être une chose facile à prouver. Toutefois, un examen plus attentif nous ramène à un problème sur la divisibilité dans le triangle de Pascal, un sujet d'étude qui s'est avéré hautement non trivial.

D'abord, nous prouvons la borne inférieure non triviale de 15/32 pour «la moitié» des entiers t. En outre, nous considérons les valeurs c_t pour t de la forme spéciale $t = ((10)^j)_2$, et montrons qu'elles sont supérieures à $1/2$ pour j assez grand. Le théorème principal, cependant, est un résultat de densité 1 concernant c_t.

Théorème. *Nous avons, pour $T \to \infty$,*

$$T - |\{t \leq T : c_t > 1/2\}| = O\left(\frac{T}{\sqrt{\log T}}\right),$$

c'est-à-dire, $c_t > 1/2$ est vrai pour t dans un sous-ensemble de $\mathbb{N}$ de densité asymptotique égale à 1.

La preuve repose sur un argument concernant la concentration des valeurs autour de la valeur attendue, c'est-à-dire, sur la méthode du deuxième moment. Nous appliquons cette méthode à la suite $(X_\lambda)_\lambda$ de variables aléatoires $X_\lambda : t \mapsto c_t$, où t est contenu dans l'intervalle fini $[2^\lambda, 2^{\lambda+1})$. La valeur moyenne est facilement déterminée et se situe bien au-dessus de $1/2$. Cependant, la suite des moments d'ordre 2 s'avère être très difficile à expliciter, car elle est caractérisée comme la *diagonale* d'une fonction rationnelle à trois variables. Nous extrayons le coefficient $[x^n y^n z^n]$ en utilisant la formule intégrale de Cauchy multivariée et la méthode du point col.

Le **chapitre 4** est une version presque inchangée de mon article "Piatetski-Shapiro sequences via Beatty sequences", qui est accepté pour publication dans Acta Arithmetica. Nous approchons les suites de la forme $(\lfloor n^c \rfloor)_n$ localement par des suites de Beatty $(\lfloor n\alpha + \beta \rfloor)_n$ (ce qui est essentiellement l'approximation de Taylor) afin de prouver le théorème suivant.

Théorème. *Supposons que f est une fonction réelle deux fois continûment dérivable sur $\mathbb{R}^+$ de telle sorte que $f, f', f'' > 0$ et qu'il existe $c_1 \geq 1/2$ et $c_2 > 0$ tel que pour $0 < x \leq y \leq 2x$ nous avons $c_1 f''(x) \leq f''(y) \leq c_2 f''(x)$. Soit $A_0 \geq 2$ de sorte que $f'(A_0) \geq 1$. Il existe une constante $C = C(f)$ telle que pour toute fonction arithmétique φ bornée par 1, pour tous les entiers $A \geq A_0$ et pour tout $z > 0$ nous avons*

$$\frac{1}{A}\left|\sum_{A<n\leq 2A} \varphi\left(\lfloor f(n) \rfloor\right) - \sum_{f(A)<m\leq f(2A)} \varphi(m)\left(f^{-1}\right)'(m)\right|$$

$$\leq C\left(\frac{f''(A)}{f'(A)^2}z^2 + f'(A)(\log A)^3 J(A,z)\right),$$

où

$$J(A,z) = \int_0^1 \sup_{f(A)<x\leq f(2A)} \frac{1}{z}\left|\sum_{x<m\leq x+z} \varphi(m)\,\mathrm{e}\,(m\vartheta)\right| \mathrm{d}\vartheta. \tag{3}$$

En particulier, ce théorème donne une condition suffisante pour qu'une fonction arithmétique φ évaluée à $\lfloor n^c \rfloor$ se comporte «comme prévu ». En utilisant ce résultat, nous étudions le comportement de la suite Thue-Morse $(0,1,1,0,1,0,0,1,1,0,0,1,0,1,1,0,\ldots)$ sur $\lfloor n^c \rfloor$, améliorant la borne $1,4$ obtenu dans [40] à $1,42$.

Théorème. *Pour $1 < c \leq 1.42$, il existe $\eta > \max\{0,(7-5c)/9\}$ et C tels que pour tout $N \geq 2$*

$$\frac{1}{N}\left|\sum_{1\leq n\leq N} (-1)^{s_2(\lfloor n^c \rfloor)}\right| \leq CN^{-\eta}.$$

Nous donnons deux autres applications du résultat principal. La deuxième application concerne la distribution conjointe des sommes des chiffres de $\lfloor n^c \rfloor$ dans les classes de congruence pour les différentes bases. Pour prouver ce résultat, nous utilisons à nouveau les coefficients de Fourier $F_\lambda(h,\vartheta)$. Nous notons que cette méthode donne également une solution alternative au premier problème de Gelfond (cas $c = 1$).

La troisième application traite la distribution de la somme de chiffres de Zeckendorf de $\lfloor n^c \rfloor$ dans les classes de congruence. Dans les deux cas, on obtient un exposant c non trivial.

Le **chapitre 5** concerne la distribution conjointe des fonctions s_q et Z dans les classes de congruence. Comme nous l'avons noté plus haut, la distribution conjointe de s_{q_1} et s_{q_2} a déjà été étudiée et Kim [33] a obtenu un résultat quantitatif d'indépendance. La distribution conjointe des fonctions s_q et Z a également été étudiée (voir Coquet, Rhin et Toffin [13]), mais seulement d'une manière non quantitative. Le but de ce chapitre est de prouver le théorème suivant.

Théorème. *Soit $q \geq 2$ entier et ϑ, β des nombres réels tels que $\beta \notin \mathbb{Z}$. Alors on a*

$$\sum_{n<N} \mathrm{e}\,(\vartheta s_q(n) + \beta Z(n)) = O\left(N^{1-\eta}\right)$$

pour un $\eta > 0$.

Pour prouver ce théorème, nous voulons caractériser les entiers commençant par une suite donnée de chiffres en $\{0,1\}$ dans la représentation Zeckendorf. Le problème correspondant pour la représentation en base q est facile : la représentation de n dans la base q commence par une certaine suite $(a_0,\ldots,a_{k-1})$ de chiffres si et seulement si n se trouve dans une certaine classe de congruence modulo q^k. Ce constat nous permet d'utiliser avantageusement la transformée de Fourier discrète dans le cadre des problèmes Gelfond, par exemple. Afin d'obtenir un énoncé analogue pour le cas de numération de Zeckendorf, nous utilisons le fait que les nombres entiers n tels que les premiers chiffres de la représentation de Zeckendorf sont fixés peuvent être caractérisés par la condition $n\varphi \in I + \mathbb{Z}$, où $\varphi = (\sqrt{5}+1)/2$ et I est un intervalle correspondant aux premiers chiffres donnés. En exploitant cette propriété, nous obtenons un résultat analogue à celui que fournit la transformée de Fourier inverse (voir la proposition 5.4), ce qui nous permet de démontrer le théorème.

0.3 Introduction in English

We begin this work by introducing the basic notions that are used in Chapters 1 to 5. The central concept in this thesis is that of a *digital representation* of an integer n, also known as a *numeration system*. In a very general setting, such a system is just an injective map from the nonnegative integers to a set of sequences of *digits*.

The most well-known numeration system is the decimal system, used throughout everyday life by a substantial part of the world's population. Slightly less well-known, but still important, is the generalization of this number system to arbitrary bases q, where $q \geq 2$ is an integer: every nonnegative integer n can be written in a unique way in the form

$$n = \sum_{i \geq 0} \varepsilon_i q^i,$$

where $\varepsilon_i \in \{0, \ldots, q-1\}$ and $\varepsilon_i = 0$ for all but finitely many i. In particular the case that $q = 2$ has massively gained importance to humans in the last century, which is due to the invention of digital computing machines. The notion of the base-q representation of an integer will be present in each chapter of this thesis. There is another system that we will be concerned with at several places, the *Zeckendorf representation* of an integer n. It is based on Zeckendorf's theorem, stating that each positive integer n can be written in a unique way as a sum of Fibonacci numbers, given by $1, 2, 3, 5, 8, 13, \ldots$, where it is forbidden to take two adjacent Fibonacci numbers. (It is easy to see that the latter condition is necessary, since if both F_k and F_{k+1} occur in a representation of n as a sum of Fibonacci numbers, where k ist maximal, taking F_{k+2} instead yields another representation.) In other words, each nonnegative integer n can be written in a unique way as

$$n = \sum_{k \geq 0} \varepsilon_k F_k,$$

where $\varepsilon_k \in \{0, 1\}$ and $\varepsilon_k = 0$ for all but finitely many k, and if $\varepsilon_k = 1$, then $\varepsilon_{k+1} = 0$. This system is a special case of the so-called Ostrowski numeration system, which is based on the continued fraction expansion of a real number. The case of the Zeckendorf representation is obtained by taking the Ostrowski numeration with respect to the golden ratio φ. However, we will not be concerned with the general case. We only note that some of the proofs of our results concerning the Zeckendorf representation could be adapted to the more general case of Ostrowski numeration, yielding analogous results. Based on the above digital representations, we can define sum-of-digits functions: for any integer $q \geq 2$ let s_q be the function that adds up the digits of n in the q-ary representation. Moreover, let Z be the function returning the number of summands in the Zeckendorf representation of n.

In the last few years, the sum-of-digits function s_q has attracted some interest and great progress has been made on the so-called Gelfond problems. These problems are proposed at the end of the article [29] by Gelfond and roughly state the following:

1. Study the joint distribution in residue classes of sum-of-digits functions in different bases.

2. Find the number of prime numbers $p \leq x$ such that $s_q(p) \equiv \ell \bmod m$.

3. For any polynomial P such that $P(n) \in \mathbb{N}$ for $n \in \mathbb{N}$, study the distribution of $s_q(P(n))$ in residue classes.

The first of these problems has been completely solved by Kim [33]. The second problem and the special case $P(n) = n^2$ of the third one have been solved by Mauduit and Rivat [41, 42].

As a generalization of the sum-of-digits function s_q in base q and the corresponding function $n \mapsto \mathrm{e}(\vartheta s_q(n))$ (where $\mathrm{e}(x) = \exp(2\pi i x)$) we define the terms "$q$-additive function" and "q-multiplicative function" as follows. An arithmetic function f is called *q-additive* if there are functions $f_k : \{0, \dots, q-1\} \to \mathbb{C}$ such that $f_k(0) = 0$ and

$$f\left(\sum_{k\geq 0} a_k q^k\right) = \sum_{k\geq 0} f_k(a_k),$$

where $a_k \in \{0, \dots, q-1\}$ and equal to zero for all but finitely many k. This condition is equivalent to the requirement that $f(q^k n + b) = f(q^k n) + f(b)$ whenever k, n are nonnegative integers such that $0 \leq b < q^k$, which is the definition given in Chapter 1. Analogously, g is called *q-multiplicative* if there are functions $g_k : \{0, \dots, q-1\} \to \mathbb{C}$ such that $g_k(0) = 1$ and

$$g\left(\sum_{k\geq 0} a_k q^k\right) = \prod_{k\geq 0} g_k(a_k),$$

where $a_k \in \{0, \dots, q-1\}$ and equal to zero for almost all k.

For the case that $q = 2$, the term q-additive function has been defined in Bellman and Shapiro [5] (where these functions are called *dyadically additive*). Many authors have since studied q-additive and q-multiplicative functions, for a collection of references to the literature we refer to [40]. The sum-of-digits function arises from the general definition by setting $f_k(b) = b$.

The purpose of **Chapter 1** is twofold. First, it is intended to be an introductory chapter. We review some previously proven results on q-additive and q-multiplicative functions and also provide a characterization of the behaviour of the mean value of q-additive functions. The principal reference to the literature for this part is Delange [16].

More importantly, we want to show that the discrete Fourier transform is a valuable tool in the study of general q-multiplicative functions. Using the Fourier transform in the context of the sum-of-digits function is one of the fundamental techniques in the two articles by Mauduit and Rivat cited above, and we show that this technique is useful also in the more general context. For $\lambda \geq 0$ and $h \in \mathbb{Z}$ the discrete Fourier coefficients for the q^λ-periodic function $u \mapsto \mathrm{e}\left(\vartheta s_q(u \bmod q^\lambda)\right)$ have the form

$$F_\lambda(h, \vartheta) = \frac{1}{q^\lambda} \sum_{u < q^\lambda} \mathrm{e}\left(\vartheta s_q(u) - huq^{-\lambda}\right).$$

Using Fourier coefficients for general q-multiplicative functions, we reprove in a simpler way a theorem of Coquet [10] concerning the relation of the *Fourier-Bohr spectrum* of a

q-multiplicative function g and a certain notion of pseudorandomness. The Fourier-Bohr spectrum of an arithmetic function g is the set of $\beta \in [0,1)$ such that

$$\limsup_{N\to\infty} \frac{1}{N} \left| \sum_{n<N} g(n)\, \mathrm{e}\,(\beta n) \right| > 0.$$

An arithmetic function g is called pseudorandom (in the sense of Bertrandias) if the (auto-) correlation

$$\gamma_t = \lim_{N\to\infty} \frac{1}{N} \sum_{n<N} g(n+t)\overline{g(n)}$$

exists for all $t \geq 0$ and, moreover, the quadratic mean of γ_t equals 0. The critical part of the abovementioned theorem by Coquet is the following theorem.

Theorem (Coquet)**.** *Let f be a q-additive function and $g(n) = \mathrm{e}\,(f(n))$. If the Fourier-Bohr spectrum of g is empty, then g is pseudorandom.*

Our proof of this statement is an application of van der Corput's inequality, followed by a discrete Fourier transform. This is the combination that has proven to be successful in [41, 42]. We also prove a related result for the arithmetic function $n \mapsto \mathrm{e}\,(\vartheta Z(n))$.

Theorem. *Let $\vartheta \in \mathbb{R} \setminus \mathbb{Z}$. Then the function $n \mapsto \mathrm{e}\,(\vartheta Z(n))$ is pseudorandom.*

In **Chapter 2** we take a closer look at the autocorrelation $\gamma_t(\vartheta)$ of the q-multiplicative function $n \mapsto \mathrm{e}\,(\vartheta s_q(n))$. This sequence is an example of a so-called q-regular sequence (see Drmota and Grabner [21]) and we have the following representation: if $(a_\nu \cdots a_0)_q$ is the q-ary representation of t, then

$$\gamma_t(\vartheta) = (1\,,0)\, A(a_0) \cdots A(a_\nu) \begin{pmatrix} 1 \\ u \end{pmatrix}$$

for some $u \in \mathbb{C}$ and 2×2-matrices $A(0)$ to $A(q-1)$. Using this formula, we prove that the correlations for the sum-of-digits function satisfy a symmetry property with respect to *digit reversal*: for any nonnegative integer t, let t^R be the integer obtained by writing the digits in base q in reverse order. In [47] we prove the following theorem:

Theorem (Morgenbesser and Spiegelhofer)**.** *For $q \geq 2$, $\vartheta \in \mathbb{R}$ and $t \geq 0$ set*

$$\gamma_t(\vartheta) = \lim_{N\to\infty} \frac{1}{N} \sum_{n<N} e(\vartheta(s_q(n+t) - s_q(n)).$$

Then we have

$$\gamma_t(\vartheta) = \gamma_{t^R}(\vartheta),$$

where the reflection t^R has to be taken with respect to the base q.

This result is unexpected insofar as the sum of digits of $n+t$ seems to be unrelated to the sum of digits of $n + t^R$. The proof of this statement is not long as soon as an adequate

induction hypothesis has been found.[3] Finding the induction hypothesis is the non-obvious ingredient of the proof, whereas the rest is basic linear algebra.

In the special case that $q = 2$ we establish a more general theorem.

Theorem. *Let α and β be complex numbers and $(z_n)_{n\geq 0}$ be a sequence satisfying the recurrence*

$$z_{2t} = z_t \quad \text{and} \quad z_{2t+1} = \alpha z_t + \beta z_{t+1}$$

for all $t \geq 1$. Then $z_{t^R} = z_t$, where the digit reversal is with respect to base 2.

Again the proof is by induction and the only tricky part is to find the induction hypothesis.

In **Chapter 3** we continue to study the relation of the sum of digits of n and $n + t$. It is believed [14] that for each t, the set of n such that $s_2(n + t) \geq s_2(n)$ has asymptotic density $c_t > 1/2$. At first sight, this might seem to be an easy thing to prove. However, a closer look suggests the opposite – for example, this problem can be translated to a problem on divisibility in Pascal's triangle, a subject that has proven to be highly nontrivial to handle.

First we prove the nontrivial lower bound 15/32 for "half of" the integers t. Moreover, we consider the values c_t for t of the special form $t = ((10)^j)_2$ and show that they are greater than $1/2$ for j large enough. The main theorem, however, is a density 1-result concerning c_t.

Theorem. *We have, as $T \to \infty$,*

$$T - |\{t \leq T : c_t > 1/2\}| = O\left(\frac{T}{\sqrt{\log T}}\right),$$

that is, $c_t > 1/2$ holds for t in a subset of $\mathbb{N}$ of asymptotic density 1.

The proof is by an argument on the concentration of the values around the expected value, that is, by the second moment method.

We apply this method to the sequence $(X_\lambda)_\lambda$ of random variables $X_\lambda : t \mapsto c_t$, where t is contained in the finite interval $[2^\lambda, 2^{\lambda+1})$. The mean value is easily determined and lies well above $1/2$. However, the sequence of second moments turns out to be quite elusive, being characterized as the *diagonal* of a rational function in three indeterminates. We extract the coefficient $[x^n y^n z^n]$ using the multivariate Cauchy integral formula and the saddle point method.

Chapter 4 is an almost unchanged version of my paper "Piatetski-Shapiro sequences via Beatty sequences", which is accepted for publication in Acta Arithmetica. (Some of the lemmas from the previous chapters appear a second time here, which makes this chapter more self-contained.) We approximate sequences of the form $(\lfloor n^c \rfloor)_n$ locally by Beatty sequences $(\lfloor n\alpha + \beta \rfloor)_n$ (which is basically Taylor approximation) in order to prove the following theorem.

Theorem. *Assume that f is a two times continuously differentiable real valued function on $\mathbb{R}^+$ such that $f, f', f'' > 0$ and that there exist $c_1 \geq 1/2$ and $c_2 > 0$ such that for $0 < x \leq y \leq 2x$ we have $c_1 f''(x) \leq f''(y) \leq c_2 f''(x)$. Let $A_0 \geq 2$ be such that $f'(A_0) \geq 1$.*

[3] The author came up with a proof for the case that $q = 2$, whereas the critical step for the general case, which is Proposition 2.10, is due to J. Morgenbesser.

There exists a constant $C = C(f)$ such that for all complex valued arithmetic functions φ bounded by 1, for all integers $A \geq A_0$ and for all $z > 0$ we have

$$\frac{1}{A}\left|\sum_{A<n\leq 2A}\varphi\left(\lfloor f(n)\rfloor\right) - \sum_{f(A)<m\leq f(2A)}\varphi(m)\left(f^{-1}\right)'(m)\right| \leq C\left(\frac{f''(A)}{f'(A)^2}z^2 + f'(A)(\log A)^3 J(A,z)\right), \quad (4)$$

where

$$J(A,z) = \int_0^1 \sup_{f(A)<x\leq f(2A)} \frac{1}{z}\left|\sum_{x<m\leq x+z}\varphi(m)\,\mathrm{e}\left(m\vartheta\right)\right| \mathrm{d}\vartheta. \quad (5)$$

In particular, this theorem gives a sufficient condition for an arithmetic function φ evaluated at $\lfloor n^c\rfloor$ to "behave as expected". Using this result, we study the behaviour of the Thue-Morse sequence $(0,1,1,0,1,0,0,1,1,0,0,1,0,1,1,0,\ldots)$ on $\lfloor n^c\rfloor$, pushing the limit 1.4 obtained in [40] to 1.42:

Theorem. *For $1 < c \leq 1.42$ there exist $\eta > \max\{0,(7-5c)/9\}$ and C such that for all $N \geq 2$*

$$\frac{1}{N}\left|\sum_{1\leq n\leq N}(-1)^{s_2(\lfloor n^c\rfloor)}\right| \leq CN^{-\eta}.$$

We give two more applications of the main result. The second application concerns the joint distribution of sum-of-digits of $\lfloor n^c\rfloor$ in residue classes with respect to different bases. In order to prove this result, we use again the Fourier coefficients $F_\lambda(h,\vartheta)$. We note that this method also gives an alternative solution to the first Gelfond problem, which is the case that $c = 1$.

The third application treats the distribution of the Zeckendorf sum of digits of $\lfloor n^c\rfloor$ in residue classes. In both cases we obtain a nontrivial exponent c.

Chapter 5 is concerned with the joint distribution of the functions s_q and Z in residue classes. As we noted above, the joint distribution of s_{q_1} and s_{q_2} has already been studied and Kim [33] obtained a quantitative independence result. The joint distribution of s_q and Z has also been studied (see Coquet, Rhin and Toffin [13]), but only in a non-quantitative manner. The purpose of this chapter is to prove the following theorem.

Theorem. *Let $q \geq 2$ be an integer and ϑ, β be real numbers such that $\beta \notin \mathbb{Z}$. Then*

$$\sum_{n<N}\mathrm{e}\left(\vartheta s_q(n) + \beta Z(n)\right) = O\left(N^{1-\eta}\right)$$

for some $\eta > 0$.

In order to prove this, we want to characterize integers starting with with a given sequence of digits in $\{0,1\}$ in the Zeckendorf expansion. The corresponding problem for the base-q representation is easy: the representation of n in base q starts with a certain sequence $(a_0,\ldots,a_{k-1})$ of digits if and only if n lies in a certain residue class modulo q^k. This fact

makes it possible to advantageously use the discrete Fourier transform in the context of the Gelfond problems, for example. In order to obtain an analogous statement for the Zeckendorf case, we use the fact that integers n such that the first digits in the Zeckendorf representation are fixed can be characterized by the condition $n\varphi \in I + \mathbb{Z}$, where $\varphi = (\sqrt{5}+1)/2$ and I is an interval corresponding to the given initial digits. Exploiting this property, we obtain a result that is analogous to the inverse discrete Fourier transform (see Proposition 5.4), which enables us to prove the theorem.

Chapter 1

q-additive and q-multiplicative functions

This chapter is concerned with two types of functions that are closely related to each other, namely q-additive and q-multiplicative functions. These kinds of functions have been studied by many authors from different points of view. The sum-of-digits function in base q is the most well-known example of a q-additive function.

To get started, we investigate criteria for the existence of the mean value of q-multiplicative functions. This has been done by Delange [16] and Kim [33]. We revisit some of the results contained in these papers.

However, the central aim of this chapter is the introduction of the discrete Fourier transform into the theory of q-additive and q-multiplicative functions. Mauduit and Rivat [41, 42] have introduced this technique in order to solve several of the so-called Gelfond problems, which are concerned with the sum-of-digits function. We show that Fourier coefficients are a useful tool also in the study of this more general kind of functions, thereby reproving a statement obtained by Coquet [10] concerning the relation of pseudorandomness and the Fourier-Bohr spectrum of a q-multiplicative function. We also generalize this result to the Zeckendorf sum-of-digits function.

The last section deals with q-multiplicative functions in different bases. We reprove a statement obtained by Coquet [10] by applying the Fourier analytic method.

1.1 Introduction and basic definitions

Throughout this chapter, we will use the notation $\|x\|$ to denote the distance of x to the nearest integer, and $\{x\}$ to denote the fractional part of x, defined by $\{x\} = x - \lfloor x \rfloor$. The function $\mathrm{e}(x)$ is the exponential function $\mathrm{e}(x) = \exp(2\pi i x)$. We also define $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$.

We begin this work with the definition of the sum-of-digits function in base q, defined as follows. Each nonnegative integer n admits a unique representation

$$n = \sum_{i \geq 0} \varepsilon_i q^i,$$

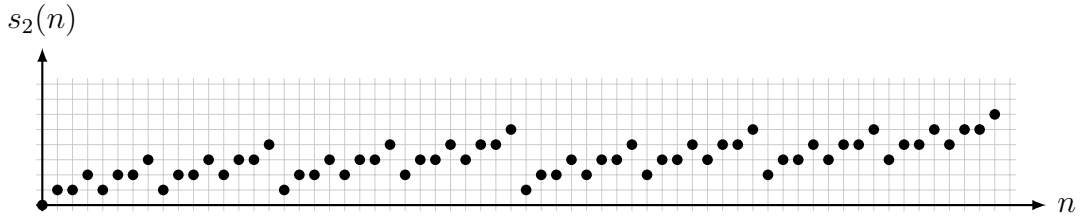
where $\varepsilon_i \in \{0, \dots, q-1\}$ for all $i \geq 0$ and for all but finitely many i we have $\varepsilon_i = 0$. We may therefore define functions $\varepsilon_i : \mathbb{N} \to \{0, \dots, q-1\}$ in such a way that

$$n = \sum_{i \geq 0} \varepsilon_i(n) q^i$$

for all $n \geq 0$. The *sum of digits of n in base q* is defined by

$$s_q(n) = \sum_{i \geq 0} \varepsilon_i(n).$$

The function s_q is called the *sum-of-digits function in base q*. We note that $s_q(qa+b) = s_q(a) + s_q(b)$ for $b < q$, which causes the "fractal" behaviour of the sum-of-digits function visible in the following plot of some values of the function s_2.



It is interesting to note that the sum of digits $s_q(n)$ of n is the least nonnegative integer k such that n can be written as a sum of k powers of q, where each power may occur more than $q-1$ times. To prove this statement, let $\mathcal{A} = \{(\eta_i)_{i \geq 0} : \eta_i = 0 \text{ for almost all } i \geq 0\}$. It is not difficult to see that for $\eta \in \mathcal{A}$ such that $\sum_{i \geq 0} \eta_i q^i = n$ and $\eta_j \geq q$ for some j there is $\mu \in \mathcal{A}$ such that $\sum_{i \geq 0} \mu_i q^i = n$ and $\sum_{i \geq 0} \mu_i < \sum_{i \geq 0} \eta_i$. We simply set $\mu_j = \eta_j \bmod q$, $\mu_{j+1} = \eta_{j+1} + \lfloor \eta_j / q \rfloor$ and $\mu_k = \eta_k$ for $k \notin \{j, j+1\}$. Iterating this procedure and applying the method of infinite descent we see that there exists a finite sequence $(\eta, \mu^1, \dots, \mu^k)$ such that $\sum_{j \geq 1} \mu_j^k < \dots < \sum_{j \geq 1} \mu_j^1 < \sum_{j \geq 1} \eta_j$ and $\sum_{j \geq 1} \mu_j^k q^j = \dots = \sum_{j \geq 1} \mu_j^k q^j = \sum_{j \geq 1} \eta_j q^j = n$ and $\mu_j^k < q$ for all $j \geq 1$. By uniqueness of the digital representation we have $\mu_j^k = \varepsilon_j(n)$ for all j and the statement follows.

The definition of the sum-of-digits function can be generalized by introducing weights, leading to the term q-additive function.

Definition 1.1. Let $q \geq 2$ be an integer. An arithmetic function $f : \mathbb{N} \to \mathbb{C}$ is called *q-additive* if

$$f(q^k n + b) = f(q^k n) + f(b)$$

for all integers $k, n > 0$ and $0 \leq b < q^k$. Moreover, f is called *completely q-additive* if $f(q^k n + b) = f(n) + f(b)$ for k, n, b satisfying the same restrictions.

For any q-additive function f we have $f(q) = f(q^1 \cdot 1 + 0) = f(q) + f(0)$, therefore $f(0) = 0$.

Let f be a q-additive function. Then f is a sum of functions on the digits: Let $\varepsilon_k \in \{0, \dots, q-1\}$ for all $k \geq 0$ and $\varepsilon_k \neq 0$ for only finitely many k. Then

$$f\big(\sum_{k\geq 0}\varepsilon_k q^k\big)=\sum_{k\geq 0} f(\varepsilon_k q^k).$$

This is clear if $\varepsilon_k=0$ for all k. We prove the statement by induction on $m=\max\{k\in\mathbb{N}:\varepsilon_k\neq 0\}$. If $m=0$, then clearly $f(\varepsilon_0)=\sum_{k\geq 0} f(\varepsilon_k q^k)$. Assume therefore that $m>0$. If $\varepsilon_k=0$ for $0\leq k<m$, the statement is obvious. Let $n=\max\{k\in\{0,\ldots,m-1\}:\varepsilon_k\neq 0\}$. Then

$$\begin{aligned} f\big(\sum_{k\geq 0}\varepsilon_k q^k\big)=f\big(\varepsilon_m q^m+\sum_{0\leq k\leq n}\varepsilon_k q^k\big)&=f(\varepsilon_m q^m)+f\big(\sum_{0\leq k\leq n}\varepsilon_k q^k\big)\\ &=f(\varepsilon_m q^m)+\sum_{0\leq k\leq n} f(\varepsilon_k q^k)=\sum_{k\geq 0} f(\varepsilon_k q^k).\end{aligned}$$

We get therefore uniquely determined functions $f_k:\{0,\ldots,q-1\}\to\mathbb{C}$ such that

$$f\big(\sum_{0\leq k\leq m}\varepsilon_k q^k\big)=\sum_{0\leq k\leq m} f_k(\varepsilon_k)$$

for all $m\geq 0$ and all $\varepsilon_k\in\{0,\ldots,q-1\}$, namely the functions $f_k(b)=f(bq^k)$. In particular we have $f_k(0)=0$ for all $k\geq 0$.

The definition of a q-additive function is to be compared to the term of a (usual) additive function. Such a function f satisfies the property that if $(a,b)=1$, we have $f(ab)=f(a)+f(b)$. If we write $a=\prod_p p^{\eta_p}$ and $b=\prod_p p^{\mu_p}$, the condition $(a,b)=1$ is the same as to claim that $\eta_p\neq 0\Rightarrow\mu_p=0$.

We can formulate the definition of a q-additive function in a way that looks very similar. Let f be q-additive and write $a=\sum_{k\geq 0}\delta_k q^k$ and $b=\sum_{k\geq 0}\varepsilon_k q^k$. The statement that f is q-additive is equivalent to condition that if $\delta_k\neq 0\Rightarrow\varepsilon_k=0$, then $f(a+b)=f(a)+f(b)$.

Definition 1.2. Let $q\geq 2$ be an integer. A function $g:\mathbb{N}\to\mathbb{C}$ is called q-multiplicative if $g(0)=1$ and $g(q^k n+b)=g(q^k n)g(b)$ for $k,n>0$ and $0\leq b<q^k$. A *completely q-multiplicative* function is defined in the obvious way.

Let ϑ and β be real numbers and $q\geq 2$ an integer. The function

$$n\mapsto \mathrm{e}\left(\vartheta s_q(n)-\beta n\right)$$

is a q-multiplicative function that will appear later on.

In analogy to the case of q-additive functions there are uniquely determined functions $g_k:\{0,\ldots,q-1\}\to\mathbb{C}$ such that

$$g\big(\sum_{0\leq k\leq m}\varepsilon_k q^k\big)=\prod_{0\leq k\leq m} g_k(\varepsilon_k)$$

for all $m\geq 0$ and $\varepsilon_k\in\{0,\ldots,q-1\}$. We have $g_k(b)=g(bq^k)$, in particular we have $g_k(0)=1$.

Many authors have studied q-multiplicative functions from various points of view, see for example [40] and the references contained therein.

1.2 A single base

1.2.1 Criteria for the existence of a mean value

The aim of this section is to study the mean value

$$\lim_{N\to\infty}\frac{1}{N}S(N) \tag{1.1}$$

of a q-multiplicative function g, where $S(N)$ denotes the partial sum

$$S(N)=\sum_{n<N}g(n).$$

Delange [16] has given necessary conditions for the case that the mean value (1.1) exists and is nonzero. Kim [34] gave a characterization of the case that the mean value is zero. We will summarize these results in Theorem 1.5, without adding any new ideas to it.

Moreover, using this theorem, we investigate the mean value of q-multiplicative functions of the form $n\mapsto \mathrm{e}\left(\vartheta f(n)\right)$, where f is q-additive, leading to Theorem 1.6. This result, which is a characterization of the mean value of such q-multiplicative functions, does not seem to be stated elsewhere in the literature, however.

Delange [16] already observed that if $q^{-k}S\left(q^k\right)$ converges, then g has a mean value, that is, the limit in (1.1) exists. We prove this result in a different way.

Proposition 1.3. *Let g be a q-multiplicative function bounded by 1 and suppose that the limit*

$$\lim_{k\to\infty}\frac{1}{q^k}S\left(q^k\right)=\lim_{k\to\infty}\frac{1}{q^k}\sum_{n<q^k}g(n)$$

exists. Then g has a mean value, that is, the limit

$$\lim_{N\to\infty}\frac{1}{N}S(N)$$

exists.

Remark. The limit

$$\lim_{k\to\infty}q^{-k}\left|S\left(q^k\right)\right|$$

always exists if g is a q-multiplicative function g that is absolutely bounded by 1.

To see this, we assume that $b\le q$. Then we have

$$S\left(bq^k\right)=\sum_{n=0}^{bq^k-1}g(n)=\sum_{m=0}^{b-1}\sum_{n<q^k}g(n+mq^k)=\sum_{m=0}^{b-1}g_k(m)\sum_{n=0}^{q^k-1}g(n)=S\left(q^k\right)\sum_{m=0}^{b-1}g_k(m).$$

In particular, we get for all $k\ge 0$ the product representation

$$\sum_{n<q^k}g(n)=\prod_{i<k}\sum_{0\le b<q}g\left(bq^i\right) \tag{1.2}$$

and consequently we have

$$\lim_{k\to\infty}\frac{1}{q^k}\left|S\left(q^k\right)\right| = \lim_{k\to\infty}\prod_{i<k}\frac{1}{q}\left|\sum_{0\le b<q} g\left(bq^i\right)\right|.$$

Since $|g(bq^i)| \le 1$, the factors of the product on the right hand side are nonnegative real numbers bounded by 1, therefore the limit on the right hand side exists. By (1.2) the limit on the left exists and both sides are equal. In the proof of Proposition 1.3 we will use the following Lemma that is (also) due to H. Delange [16].

Lemma 1.4. *Let $z_1, \dots, z_{q-1}$ be complex numbers such that $|z_j| \le 1$ for $1 \le j < q$. Then*

$$\left|\frac{1}{q}(1 + z_1 + \cdots + z_{q-1})\right| \le 1 - \frac{1}{2q}\max_{1\le j<q}\left(1 - \operatorname{Re} z_j\right).$$

Proof of Proposition 1.3. Let M be the limit $\lim_{k\to\infty} q^{-k}S\left(q^k\right)$. For $k \ge 0$ we set

$$N_k = \sum_{i\ge k}\varepsilon_i(N)q^i.$$

Let $N \ge 1$ and $L = L(N)$ the length of the q-adic expansion of N, that is, $L = \max\{k : \varepsilon_k(N) \ne 0\}$. By induction on L one easily shows the validity of the representation

$$\sum_{n<N} g(n) = \sum_{k\le L}\sum_{n<\varepsilon_{L-k}(N)q^{L-k}} g(n + N_{L-k+1}).$$

By a change of variables we further get

$$\begin{aligned} S(N) &= \sum_{k\le L}\sum_{n<\varepsilon_k(N)q^k} g(n + N_{k+1}) \\ &= \sum_{k\le L} g(N_{k+1}) \sum_{b<\varepsilon_k(N)}\sum_{n<q^k} g(n + bq^\lambda) \\ &= \sum_{k\le L} g(N_{k+1})S\left(q^k\right)\sum_{b<\varepsilon_k(N)} g_k(b). \end{aligned} \tag{1.3}$$

We treat the case that $M \ne 0$ first. We show that for all $b < q$ we have $g_k(b) \to 1$ as $k \to \infty$ by contradiction. Assume the opposite, then for some $b_0 < q$ and some $\varepsilon > 0$ we have $|g_k(b_0) - 1| > \varepsilon$ infinitely often. Since $\operatorname{Re} g_k(b_0) < 1$ for infinitely many k, we obtain from Lemma 1.4 and $g(0) = 1$ that the values

$$\frac{1}{q}\sum_{b<q} g\left(bq^k\right)$$

do not converge to 1 as $k \to \infty$. It follows that the values

$$q^{-\lambda}S\left(q^\lambda\right) = \prod_{k<\lambda}\frac{1}{q}\sum_{b<q} g_k(b).$$

tend to 0 as $\lambda \to \infty$. Therefore $M = 0$, which leads to a contradiction. Consequently $g_k(b) \to 1$ as $k \to \infty$.

From this property it follows that

$$\sum_{b<\varepsilon_k(N)} g_k(b) - \varepsilon_k(N)$$

converges to 0 uniformly in N as $k \to \infty$. Moreover

$$\frac{1}{Mq^k} S\left(q^k\right) \to 1$$

as $k \to \infty$. Finally for each $t \in \mathbb{N}$ and all $\varepsilon > 0$ we have

$$|g(N_k) - 1| < \varepsilon$$

for all sufficiently large k and all $N \leq q^{t+k}$. Combining these observations, we see after a short calculation that for given t we have

$$\left| \frac{1}{N} \sum_{k=L-t+1}^{L} g(N_{k+1}) S\left(q^k\right) \sum_{b<\varepsilon_k(N)} g_k(b) - M \right| \ll q^{-t}.$$

as $N \to \infty$. Moreover, we have the trivial estimate

$$\left| \frac{1}{N} \sum_{k \leq L-t} g(N_{k+1}) S\left(q^k\right) \sum_{b<\varepsilon_k(N)} g_k(b) \right| \leq \frac{1}{N} \sum_{k \leq L-t} q^{k+1} \ll q^{-t}$$

with an implied constant depending only on q. Choosing t properly gives the statement $N^{-1}S(N) \to M$ which we wanted to prove.

It remains to treat the case that $M = 0$. By equation (1.3) we get

$$|S(N)| \leq q \sum_{k \leq L} \left|S\left(q^k\right)\right| \ll q^{k_0} + \sum_{k \geq k_0} q^k \frac{|S\left(q^k\right)|}{q^k}.$$

Choosing k_0 in such a way that $|S\left(q^k\right) q^{-k}| < \varepsilon$ is small for all $k \geq k_0$ it follows that $\limsup_{N\to\infty}(1/N)|S(N)| \leq \varepsilon$. Hence, $\lim_{N\to\infty}(1/N)S(N) = 0$. □

As already mentioned, we are interested in criteria for a q-multiplicative function to possess a mean value and moreover, in criteria for the mean value to be nonzero. This question has been answered completely in the articles by Delange [16] and Kim [34]. The following Theorem (and its proof) is a compilation of material that can be found in these two articles.

Theorem 1.5. *Let g be a q-multiplicative function bounded by 1 and g_k its component functions given by $g_k(b) = g\left(bq^k\right)$ for $b < q$ and $k \geq 0$. For each $k \geq 0$, define*

$$\varepsilon_k = \max_{b<q} \left(1 - \operatorname{Re} g_k(b)\right),$$

$$u_k = \frac{1}{q} \sum_{b<q} \left(g_k(b) - 1\right),$$

$$S(N) = \sum_{n<N} g(n).$$

Assume that, for all $k \geq 0$,

$$\sum_{b<q} g_k(b) \neq 0.$$

Then the following properties hold:

(a) We have $\lim_{N\to\infty} \frac{1}{N} S(N) = 0$ *if and only if* $\sum_{k\geq 0} \varepsilon_k = \infty$.

(b) $(1/N)S(N)$ *converges to a nonzero limit as* $N \to \infty$ *if and only if* $\sum_{k\geq 0} u_k$ *converges.*

(c) $(1/N)S(N)$ *is divergent as* $N \to \infty$ *if and only if* $\sum_{k\geq 0} \varepsilon_k < \infty$ *and* $\sum_{k\geq 0} u_k$ *diverges.*

Note that if $\sum_{b<q} g_k(b) = 0$ for some k, the mean value of g is zero by (1.2) and Proposition 1.3.

Proof. We first note that by Proposition 1.3 the product

$$\prod_{k\geq 0}(1+u_k) = \prod_{k\geq 0} \frac{1}{q} \sum_{b<q} g_k(b)$$

is convergent if and only if g has a nonzero mean value. (Recall that an infinite product $\prod_{k=0}^{\infty} a_k$ is called *convergent* if $a_n \neq 0$ for all $n \geq n_0$ and $\prod_{k=n_0}^{n} a_k$ converges to a nonzero limit as $n \to \infty$.)

By Lemma 1.4 and the inequality $1 + x \leq e^x$ we have

$$\left| \frac{1}{q} \sum_{b<q} g_k(b) \right| \leq 1 - \frac{\varepsilon_k}{2q} \leq \exp\left(-\frac{\varepsilon_k}{2q}\right),$$

therefore

$$\left| \prod_{k<K} (1+u_k) \right| \leq \exp\left(-\frac{1}{2q} \sum_{k<K} \varepsilon_k\right)$$

and we see that the hypothesis $\sum_{k\geq 0} \varepsilon_k = \infty$ implies $M = 0$. To prove the converse of the first assertion, we may restrict ourselves to the case that $\varepsilon \to 0$ as $k \to \infty$. Let R be large enough that $\varepsilon_k < 1$ for all $k \geq R$. By the inequality

$$\left| \sum_{0\leq b<q} z_b \right| \geq \sum_{0\leq b<q} \operatorname{Re} z_b = \sum_{0\leq b<q} (1 - (1 - \operatorname{Re} z_b)) \geq q\left(1 - \max_{0\leq b<q}(1 - \operatorname{Re} z_b)\right)$$

that holds for all complex numbers $z_0, \ldots, z_{q-1}$ we have $\frac{1}{q}\left|\sum_{b<q} g_k(b)\right| \geq 1 - \varepsilon_k$. From the fact that $\frac{1}{N}\sum_{n<N} g(n)$ converges to 0 as $N \to \infty$ and the representation

$$\frac{1}{q^K} \sum_{n<q^K} g(n) = \prod_{k<K} \frac{1}{q} \sum_{b<q} g\left(bq^k\right)$$

we conclude that the products $\prod_{R\le k<K}(1-\varepsilon_k)$ tend to zero as $K\to\infty$. Since $0\le\varepsilon_k<1$, the claim follows.

We prove the second statement. By using the Cauchy-Schwarz inequality and the inequality $|z-1|^2 = 1-2\operatorname{Re} z+(\operatorname{Re} z)^2+(\operatorname{Im} z)^2 \le 2-2\operatorname{Re} z$ (that holds for all z such that $|z|\le 1$) we obtain

$$|u_k|^2 = \frac{1}{q^2}\left|\sum_{1\le b<q}\left(g\left(bq^k\right)-1\right)\right|^2 \le \frac{q-1}{q^2}\sum_{1\le b<q}\left|g\left(bq^k\right)-1\right|^2$$
$$\le \frac{2(q-1)}{q^2}\sum_{1\le b<q}\left(1-\operatorname{Re} g\left(bq^k\right)\right) \le \frac{2(q-1)^2}{q^2}\varepsilon_k,$$

therefore $\sum_{k\ge0}\varepsilon_k<\infty$ implies $\sum_{k\ge0}|u_k|^2<\infty$.

It follows that under the assumption

$$\sum_{k\ge0}\varepsilon_k<\infty \tag{1.4}$$

the product $\prod_{k\ge0}(1+u_k)$ is convergent if and only if the series $\sum_{k\ge0}u_k$ is convergent.

To prove (b), it remains to show that we may drop the assumption (1.4). But convergence of the product $\prod_{k\ge0}(1+u_k)$ implies (1.4) by (a); moreover, we have $-q\operatorname{Re} u_k = \sum_{1\le b<q}\left(1-\operatorname{Re} g\left(bq^k\right)\right)\ge q\varepsilon_k$, therefore also convergence of $\sum_{k\ge0}u_k$ implies (1.4).

The statement (c) is a simple consequence of (a) and (b). □

We apply these findings to a special class of q-multiplicative functions g, given by $n\mapsto \mathrm{e}(f(n))$, where f is q-additive. Again the results that follow are minor variations of the material presented in Delange [16] and Kim [34].

For each real number r let $\langle r\rangle = r-\left\lfloor r+\frac12\right\rfloor$. (This function can also be written as $\langle r\rangle=\psi(x+1/2)$, where $\psi(x)=\{x\}-1/2$ and gives the "signed distance to the nearest integer".) Note that $-\frac12\le\langle r\rangle<\frac12$ and $|\langle r\rangle| = \|r\|$.

Theorem 1.6. *Let f be a q-additive function. Then the limit*

$$M=\lim_{N\to\infty}\frac{1}{N}\sum_{n<N}\mathrm{e}(f(n))$$

exists if and only if

(a) $\displaystyle\sum_{b<q}\mathrm{e}\left(f\left(bq^k\right)\right)=0$ *for some* $k\ge0$ *or*

(b) $\displaystyle\sum_{k\ge0}\sum_{b<q}\left\|f\left(q^kb\right)\right\|^2=\infty$ *or*

(c) $\displaystyle\sum_{k\ge0}\sum_{b<q}\left\|f\left(q^kb\right)\right\|^2<\infty$ *and* $\displaystyle\sum_{k\ge0}\sum_{b<q}\left\langle f\left(q^kb\right)\right\rangle$ *converges.*

If (a) or (b) is satisfied we have $M=0$. If (c) holds but (a) is violated, we have $M\ne0$.

The proof requires some technical lemmas.

Lemma 1.7. *If $(x_k)_{k\geq 0}$ is a sequence of real numbers bounded by π, then $\sum_{k\geq 0}(1-\cos(x_k))$ converges if and only if $\sum_{k\geq 0} x_k^2$ does.*

Proof. We have $1-\cos(x) = 1/2x^2+O(x^4)$ as $x\to 0$, therefore the statement follows from the limit comparison test as soon as we have shown that $x_k \to 0$. Let $\sum_{k\geq 0}(1-\cos(x_k)) < \infty$ Convergence of this sum implies $\cos(x_k)\to 1$ as $k\to\infty$ and by continuity of the function $\arccos : [0,\pi]\to[-1,1]$ and the condition $|x_k|\leq \frac{1}{2}$ we get $x_k\to 0$ for $k\to\infty$. □

Lemma 1.8. *Let f be a q-additive function and $g(n) = \mathrm{e}(f(n))$. For $k\geq 0$ define*

$$\varepsilon_k = \max_{b<q}\left(1-\operatorname{Re} g\left(bq^k\right)\right).$$

Then

$$\sum_{k\geq 0}\varepsilon_k < \infty$$

if and only if

$$\sum_{k\geq 0}\sum_{b<q}\left\|f\left(bq^k\right)\right\|^2 < \infty.$$

Proof. The series $\sum\varepsilon_k$ converges if and only if for each $b\in\{0,\dots,q-1\}$ the series

$$\sum_{k\geq 0}\left(1-\cos\left(2\pi\left\|f\left(bq^k\right)\right\|\right)\right)$$

converges. By Lemma 1.7 this is the case if and only if for all $b\in\{0,\dots,q-1\}$ we have $\sum_{k\geq 0}\left\|f\left(q^kb\right)\right\|^2<\infty$, which is equivalent to $\sum_{k\geq 0}\sum_{b<q}\left\|f\left(q^kb\right)\right\|^2<\infty$.

□

Lemma 1.9. *Let $f:\mathbb{N}\to\mathbb{C}$ be a q-additive function. Then the following statements are equivalent:*

(1) The series

$$\sum_{k\geq 0}\left(q-1-\sum_{1\leq b<q} e\left(f\left(bq^k\right)\right)\right) \tag{1.5}$$

converges.

(2) The series

$$\sum_{k\geq 0}\sum_{1\leq b<q}\left\|f\left(bq^k\right)\right\|^2 \tag{1.6}$$

and

$$\sum_{k\geq 0}\left(\sum_{1\leq b<q}\left\langle f\left(bq^k\right)\right\rangle\right) \tag{1.7}$$

converge.

Proof. Convergence of (1.5) is equivalent to the convergence of

$$\sum_{k\geq 0}\left(q-1-\sum_{1\leq b<q}\cos\left(2\pi f\left(bq^k\right)\right)\right) \tag{1.8}$$

and

$$\sum_{k\geq 0}\left(\sum_{1\leq b<q}\sin\left(2\pi f\left(bq^k\right)\right)\right). \tag{1.9}$$

Since the series (1.8) has only nonnegative terms, its convergence is equivalent to the convergence of the series

$$\sum_{k\geq 0}\max_{b<q}\left(1-\cos\left(2\pi f\left(q^k b\right)\right)\right)=\sum_{k\geq 0}\varepsilon_k.$$

By Lemma 1.8 this is equivalent to (1.6).

There is a $c>0$ such that the inequality $|\sin(x)-x|\leq c\,|x|^2$ holds for all $x\in\left[-\frac{1}{2},\frac{1}{2}\right]$, therefore

$|\sin(2\pi x)-\langle x\rangle|=|\sin(2\pi\,\langle x\rangle)-\langle x\rangle|\leq 4\pi^2 c\,\|x\|^2$. Consequently

$$\left|\sum_{1\leq b<q}\sin\left(2\pi f\left(bq^k\right)\right)-\sum_{1\leq b<q}\left\langle f\left(bq^k\right)\right\rangle\right|\leq (q-1)4\pi^2 c\max_{0\leq b<q}\left\|f\left(bq^k\right)\right\|^2. \tag{1.10}$$

Now if (1.6) and (1.7) are convergent, convergence of (1.8) and (1.9) follows; Conversely, convergence of (1.8) implies convergence of (1.6) and therefore by (1.9) the series (1.7) is convergent. □

The proof of Theorem 1.6 is now an immediate consequence of Theorem 1.5 and Lemma 1.9.

1.2.2 The discrete Fourier transform and q-multiplicative functions

Let g be a q-multiplicative function bounded by 1. For every positive integer λ we define $g_\lambda(n)$ as

$$g_\lambda(n)=g(n \bmod q^\lambda),$$

where $n \bmod q^\lambda\in\{0,1,\ldots,q^\lambda-1\}$ denotes the residue class of n modulo q^λ. It is clear that g_λ is periodic with period q^λ. Note also that $g_\lambda(n)$ depends only on the digits $\varepsilon_0(n),\ldots,\varepsilon_{\lambda-1}(n)$.

We define the discrete Fourier coefficients $G_\lambda(h)$ of the q^λ-periodic functions g_λ.

Definition 1.10. Assume that g is a q-multiplicative function, $\lambda\geq 0$ and $h\in\mathbb{Z}$. We write

$$G_\lambda(h)=\frac{1}{q^\lambda}\sum_{u<q^\lambda}g(u)\,\mathrm{e}\left(-huq^{-\lambda}\right).$$

These Fourier terms were introduced by Mauduit and Rivat [41, 42] in relation to the sum-of-digits function, that is, they used them for $g(n) = e(\vartheta s_q(n))$.

By inverting the Fourier transform we obtain

$$g_\lambda(n) = \sum_{h<q^\lambda} G_\lambda(h)\, \mathrm{e}\left(hnq^{-\lambda}\right)$$

for all nonnegative integers n and by a short calculation we also get

$$g_\lambda(-n) = \sum_{h<q^\lambda} \overline{G_\lambda(-h)}\, \mathrm{e}\left(hnq^{-\lambda}\right).$$

Since the function $n \mapsto f(n)\,\mathrm{e}\left(-hnq^{-\lambda}\right)$ is q-multiplicative, we have the product representation

$$G_\lambda(h) = \frac{1}{q^\lambda} \prod_{k<\lambda} \sum_{b<q} f^{(k)}(b)\, \mathrm{e}\left(-hbq^{k-\lambda}\right). \tag{1.11}$$

In the special case that $g(n)$ is the Thue-Morse sequence, $g(n) = \mathrm{e}\,(1/2s_2(n))$, we get

$$G_\lambda(h) = \frac{1}{2^\lambda} \prod_{k<\lambda} \left(1 - \mathrm{e}\left(-h2^{k-\lambda}\right)\right) = \frac{1}{2^\lambda} \prod_{1\le k\le\lambda} \left(1 - \mathrm{e}\left(-h2^{-k}\right)\right),$$

which is an identity that we will use in chapter 4.

In order to show the usefulness of these discrete Fourier terms we present the following new observation.

Theorem 1.11. *Let g be a q-multiplicative function and $G_\lambda(h)$ the corresponding Fourier coefficients. If*

$$\sup_{h<q^\lambda} |G_\lambda(h)| = o(1), \tag{1.12}$$

then

$$\sup_{\beta\in\mathbb{R}} \left| \frac{1}{N} \sum_{n<N} g(n)\, \mathrm{e}\,(n\beta) \right| = o(1).$$

Actually we also get a quantified version. If

$$\sup_{h<q^\lambda} |G_\lambda(h)| \ll q^{-\varepsilon} \tag{1.13}$$

for some $\varepsilon > 0$, we have

$$\frac{1}{N} \sum_{n<N} g(n)\, \mathrm{e}\,(n\beta) = N^{-\eta} \tag{1.14}$$

uniformly in β, for some $\eta = \eta(\varepsilon) > 0$.

By using the product representation (1.11) and by grouping together consecutive factors it is shown by Mauduit and Rivat [41, Lemme 9] that (1.13) holds for functions $g(n) = \mathrm{e}\,(\vartheta s_q(n))$, where $(q-1)\vartheta \notin \mathbb{Z}$.

Lemma 1.12 (Mauduit, Rivat). *Let $q, \lambda \geq 2$ and h be integers and $\vartheta \in \mathbb{R}$. Then*

$$|G_{q,\lambda}(h,\vartheta)| \leq \mathrm{e}^{\pi^2/48}\, q^{-c_q \|(q-1)\vartheta\|^2 \lambda},$$

where

$$G_{q,\lambda}(h,\vartheta) = \frac{1}{q^\lambda} \sum_{u<q^\lambda} \mathrm{e}\left(\vartheta s_q(u) - huq^{-\lambda}\right)$$

and

$$c_q = \frac{\pi^2}{12 \log q}\left(1 - \frac{2}{q+1}\right).$$

We also note that Gelfond [29] arrived at an estimate of the form (1.14) for the q-multiplicative function $g(n) = \mathrm{e}\left((m/p)s_q(n)\right)$, using a different method. More precisely he showed the following statement.

Lemma 1.13 (Gelfond). *Assume that $p \geq 2, q \geq 1$ and m are integers such that $1 \leq m < p$ and $(p, q-1) = 1$. There exists $\lambda = \lambda(p,q) < 1$ such that for all real a we have*

$$\sum_{n=1}^{N} \mathrm{e}\left(na + \frac{m}{p} s_q(n)\right) \ll N^\lambda.$$

By using again a simple discrete Fourier transform we can transfer (1.14) into a statement about the distribution of the sum-of-digits function in residue classes: for $x \in \mathbb{N}$ we have

$$\begin{aligned}
&|\{n < x : n \equiv \ell \bmod m, s_q(n) \equiv a \bmod p\}| \\
&\quad = \sum_{n<x} \frac{1}{m} \sum_{k_1<m} \mathrm{e}\left(k_1 \frac{n-\ell}{m}\right) \frac{1}{p} \sum_{k_2<p} \mathrm{e}\left(k_2 \frac{s_q(n)-a}{p}\right) \\
&\quad = \frac{1}{mp} \sum_{\substack{k_1<m \\ k_2<p}} \mathrm{e}\left(-k_1\frac{\ell}{m} - k_2\frac{a}{p}\right) \sum_{n<x} \mathrm{e}\left(n\frac{k_1}{m} + \frac{k_2}{p} s_q(n)\right) \\
&\quad = \frac{x}{mp} + O\left(\frac{1}{mp} \sum_{1\leq k_1<m} \left|\sum_{n<x} \mathrm{e}\left(n\frac{k_1}{m}\right)\right| + \frac{1}{mp} \sum_{\substack{0\leq k_1<m \\ 1\leq k_2<p}} \left|\sum_{n<x} \mathrm{e}\left(n\frac{k_1}{m} + \frac{k_2}{p} s_q(n)\right)\right|\right)
\end{aligned} \tag{1.15}$$

and by using the Lemma and $\left|\sum_{n<N} \mathrm{e}\left(nk_1/m\right)\right| \leq m$ we obtain the following result.

Theorem 1.14 (Gelfond). *Let $q, p \geq 2$ be positive integers with $(p, q-1) = 1$. There exists $\lambda < 1$ such that for all integers $m \geq 2$, a, l and for all $x \geq 1$ we have*

$$|\{1 \leq n \leq x : n \equiv l \bmod m, s_q(n) \equiv a \bmod p\}| = \frac{x}{mp} + O(x^\lambda).$$

We return to Theorem 1.11, which we want to prove now. For the proof we need a series of lemmas. The inequality of van der Corput is well known, see for example [41] for a proof.

Lemma 1.15 (Van der Corput's inequality)**.** *Let I be a finite interval in $\mathbb{Z}$ and let $a_n \in \mathbb{C}$ for $n \in I$. Then*

$$\left|\sum_{n\in I} a_n\right|^2 \le \frac{|I|-1+R}{R} \sum_{0\le|r|<R} \left(1-\frac{|r|}{R}\right) \sum_{\substack{n\in I\\ n+r\in I}} a_{n+r}\overline{a_n}$$

for all integers $R \ge 1$.

Lemma 1.16. *Let $g_\lambda(n)$ a function with period q^λ. Then we have for all $r \ge 0$*

$$\frac{1}{q^\lambda} \sum_{0\le n<q^\lambda} g_\lambda(n+r)\overline{g_\lambda(n)} = \sum_{0\le h<q^\lambda} |G_\lambda(h)|^2 \,\mathrm{e}\left(hrq^{-\lambda}\right).$$

Proof. By definition we get

$$\begin{aligned}
&\frac{1}{q^\lambda} \sum_{0\le n<q^\lambda} g_\lambda(n+r)\overline{g_\lambda(n)} \\
&\quad= \frac{1}{q^\lambda} \sum_{0\le n<q^\lambda} \sum_{0\le h_1<q^\lambda} G_\lambda(h_1)\,\mathrm{e}\left(h_1(n+r)q^{-\lambda}\right) \sum_{0\le h_2<q^\lambda} \overline{G_\lambda(-h_2)}\,\mathrm{e}\left(h_2nq^{-\lambda}\right) \\
&\quad= \sum_{0\le h_1,h_2<q^\lambda} G_\lambda(h_1)\overline{G_\lambda(-h_2)}\,\mathrm{e}\left(h_1rq^{-\lambda}\right) \frac{1}{q^\lambda} \sum_{0\le n<q^\lambda} \mathrm{e}\left(n(h_1+h_2)q^{-\lambda}\right) \qquad (1.16)\\
&\quad= \sum_{0\le h<q^\lambda} |G_\lambda(h)|^2\,\mathrm{e}\left(hrq^{-\lambda}\right),
\end{aligned}$$

which is the statement of the Lemma. □

Next we prove the following lemma (compare to [42, Lemme 5], for example) on truncated q-multiplicative functions, which is a way of expressing the idea that addition of an integer r to n should only change digits at low positions in most cases.

Lemma 1.17. *Assume that g is a q-multiplicative function , where $q \ge 2$. Let $\lambda \ge 0$ and r be integers and let I be a finite interval in $\mathbb{N}$ such that $I + r \subseteq \mathbb{N}$. Then*

$$\left|\{n \in I : g(n+r)\overline{g(n)} \ne g_\lambda(n+r)\overline{g_\lambda(n)}\}\right| \le |I|\,\frac{|r|}{q^\lambda} + |r|\,.$$

Proof. It is sufficient to assume that r is nonnegative, since the other case then follows by shifting the interval I.

For a nonnegative integer n, there exist unique t and u such that $n = tq^\lambda + u$, where $u < q^\lambda$. Clearly we have $g(n) = g(tq^\lambda)g(u)$ and $g_\lambda(n) = g(u)$. If $n \equiv k \bmod q^\lambda$ for some k such that $0 \le k < q^\lambda - r$, then $g(n+r) = g(tq^\lambda)g(u+r)$ and $g_\lambda(n+r) = g(u+r)$, therefore $g(n+r)\overline{g(n)} = g_\lambda(n+r)\overline{g_\lambda(n)}$. It remains therefore to show that $\left|\{n \in I : q^\lambda - r \le n \bmod q^\lambda < q^\lambda\}\right| \le |I|\,r/q^\lambda + r$, which is not difficult. □

The following elementary exponential sum estimate will be used at several places throughout this work.

Lemma 1.18. *Let $x \in \mathbb{R}$ and $N \geq 0$. Then*

$$\left|\sum_{n<N} \mathrm{e}\left(nx\right)\right| \leq \min\left(N, \frac{1}{2\left\|x\right\|}\right),$$

where division by zero is understood to yield the value ∞.

Proof. We have

$$\sum_{n<N} \mathrm{e}\left(nx\right) = \frac{1-\mathrm{e}(Nx)}{1-\mathrm{e}(x)}.$$

Since $1/(1-\mathrm{e}(x)) = \mathrm{e}(-x/2)/(\mathrm{e}(-x/2)-\mathrm{e}(x/2)) = \mathrm{e}(-x/2)/(-2i\sin\pi x)$, we obtain

$$\left|\sum_{n<N} \mathrm{e}(nx)\right| \leq \frac{1}{|\sin \pi x|} \leq \frac{1}{2\left\|x\right\|},$$

the last inequality being due to the concavity of the sine function. □

Lemma 1.19. *Let $H \geq 1$ be an integer and R a real number. For all real numbers t we have*

$$\sum_{h<H}\left|\frac{1}{R}\sum_{r<R}\mathrm{e}\left(r\left(t+\frac{h}{H}\right)\right)\right|^2 \leq \frac{H+R-1}{R}.$$

This lemma is an immediate consequence of the analytic form of the large sieve (see [45, Theorem 3]. This form of the theorem, with the improved constant, is due to Selberg.)

Theorem 1.20 (Selberg)**.** *Let $N \geq 1, R \geq 1, M$ be integers, $\alpha_1, \ldots, \alpha_R \in \mathbb{R}$ and $a_{M+1}, \ldots, a_{M+N} \in \mathbb{C}$. Assume that $\|\alpha_r - \alpha_s\| \geq \delta$ for $r \neq s$. Then*

$$\sum_{r=1}^{R}\left|\sum_{n=M+1}^{M+N} a_n\,\mathrm{e}(n\alpha_r)\right|^2 \leq (N-1+\delta^{-1})\sum_{n=M+1}^{M+N}|a_n|^2.$$

Now we are ready to prove Theorem 1.11. Let g be a q-multiplicative function absolutely bounded by 1. Let N, R, λ be integers such that $1 \leq R \leq N$ and $\lambda \geq 0$. We choose them later. Moreover, let β be a real number. Let k be chosen such that $kq^\lambda \leq N < (k+1)q^\lambda$. We consider partial sums of $g(n)\,\mathrm{e}\left(n\beta\right)$ up to N. We have

$$\left|\sum_{n<N} g(n)\,\mathrm{e}\left(n\beta\right)\right|^2 \leq \left|\sum_{n<kq^\lambda} g(n)\,\mathrm{e}\left(n\beta\right)\right|^2 + q^{2\lambda} + 2Nq^\lambda.$$

We apply the inequality of van der Corput (Lemma 1.15) to obtain

$$\left|\sum_{n<N} g(n)\,\mathrm{e}\left(n\beta\right)\right|^2 \leq \frac{N+R}{R}\sum_{|r|<R}\left(1-\frac{|r|}{R}\right)\mathrm{e}\left(r\beta\right)\sum_{0\leq n,n+r<kq^\lambda} g(n+r)\overline{g(n)}.$$

We adjust the summation range by omitting the condition $0 \le n+r < kq^\lambda$. This introduces an error term $O(NR)$. Moreover, q-multiplicative functions g satisfy Lemma 1.17, therefore we may replace g by g_λ for the price of another error term, $O(N^2Rq^{-\lambda}+NR)$. Using (1.16) we get

$$\begin{aligned}\left|\sum_{n<kq^\lambda} g(n)\,\mathrm{e}\,(n\beta)\right|^2 &\\ \ll \frac{N}{R}\sum_{|r|<R}\left(1-\frac{|r|}{R}\right)\mathrm{e}\,(r\beta)\left(\sum_{0\le n<kq^\lambda} g_\lambda(n+r)\overline{g_\lambda(n)}+O\left(R+NRq^{-\lambda}\right)\right)&\\ \ll NR+N^2\frac{R}{q^\lambda}+\frac{N}{R}kq^\lambda\sum_{h<q^\lambda}|G_\lambda(h)|^2\sum_{|r|<R}\left(1-\frac{|r|}{R}\right)\mathrm{e}\left(r\left(\beta+\frac{h}{q^\lambda}\right)\right).&\end{aligned}$$

Note that the sum over r is a nonnegative real number. This follows from the identity

$$\sum_{|r|<R}(R-|r|)\,\mathrm{e}\,(rx)=\left|\sum_{r<R}\mathrm{e}\,(rx)\right|^2,$$

which can be proved by an elementary combinatorial argument. We use this equation and collect the error terms to get

$$\left|\frac{1}{N}\sum_{n<N} g(n)\,\mathrm{e}\,(n\beta)\right|^2 \ll \frac{q^{2\lambda}}{N^2}+\frac{q^\lambda}{N}+\frac{R}{N}+\frac{R}{q^\lambda}+\sum_{h<q^\lambda}|G_\lambda(h)|^2\left|\frac{1}{R}\sum_{r<R}\mathrm{e}\left(r\left(\beta+\frac{h}{q^\lambda}\right)\right)\right|^2. \quad (1.17)$$

Next, using Lemma 1.19 we get

$$\sum_{h<q^\lambda}|G_\lambda(h)|^2\left|\frac{1}{R}\sum_{r<R}\mathrm{e}\left(r\left(\beta+\frac{h}{q^\lambda}\right)\right)\right|^2 \le \sup_{h<q^\lambda}|G_\lambda(h)|^2\,\frac{q^\lambda+R-1}{R}. \quad (1.18)$$

Assume that $\varepsilon\in(0,1)$ and choose λ so large that $q^{-\lambda}<\varepsilon$ and

$$\sup_{h<q^\lambda}|G_\lambda(h)|^2<\varepsilon^2.$$

The existence of such a λ is guaranteed by the hypothesis (1.12) of the theorem we are proving. Choose $R\le q^\lambda$ in such a way that

$$\frac{1}{q}\frac{R}{q^\lambda}\le\varepsilon<q\frac{R}{q^\lambda}.$$

We obtain for all $N\ge q^\lambda/\varepsilon$

$$\left|\frac{1}{N}\sum_{n<N} g(n)\,\mathrm{e}\,(n\beta)\right|^2 \ll \varepsilon^2+2\varepsilon+2\varepsilon+2q\varepsilon+2q\varepsilon+2\varepsilon^2$$

with an absolute implied constant, where the first four terms come from the first four summands in (1.17) and the last two summands come from (1.18). This finishes the proof of Theorem 1.11.

1.2.3 Sequences having empty Fourier-Bohr spectrum

We begin with the definition of two notions.

Definition 1.21. Let g be an arithmetical function. The set of $\beta \in [0,1)$ satisfying

$$\limsup_{N\to\infty} \frac{1}{N} \left| \sum_{n<N} g(n)\, \mathrm{e}\,(-n\beta) \right| > 0$$

is called the *Fourier-Bohr spectrum* of f.

The function f is called *pseudorandom in the sense of Bertrandias* or simply *pseudorandom* if the autocorrelation

$$\gamma_r = \lim_{N\to\infty} \frac{1}{N} \sum_{n<N} g(n+r)\overline{g(n)}$$

exists for all $r \geq 0$ and is zero in quadratic mean, that is,

$$\lim_{R\to\infty} \frac{1}{R} \sum_{r<R} |\gamma_r|^2 = 0. \tag{1.19}$$

These terms are also defined in J. Coquet's thesis [10]. In this section we investigate the relation between these two notions, given that g is a q-multiplicative function. This connection has been established by Coquet [10]. In particular, he proved the critical implication that if a q-multiplicative function g has empty spectrum, then it is pseudorandom. We reprove this implication in a simpler way, using the discrete Fourier transform, thus confirming the usefulness of Fourier terms again.

We note that for bounded functions g the condition (1.19) is equivalent to the property that

$$\lim_{R\to\infty} \frac{1}{R} \sum_{r<R} |\gamma_r| = 0.$$

The property is sufficient since $|\gamma_r|^2 \leq K\,|\gamma_r|$ for some bound K independent of r. To prove the opposite, we use the Cauchy-Schwarz inequality to obtain

$$\left(\frac{1}{R} \sum_{r<R} |\gamma_r| \right)^2 \leq \frac{1}{R^2} \sum_{r<R} 1 \sum_{r<R} |\gamma_r|^2 = \frac{1}{R} \sum_{r<R} |\gamma_r|^2 .$$

The statement follows since $a_n \to 0$ if and only if $a_n^2 \to 0$ (for any sequence a_n).

By reformulating Theorem 1.11 of the previous section by using the above notation we have: *If g is a q-multiplicative function absolutely bounded by 1 and the Fourier coefficients $G_\lambda(h)$ converge to 0 uniformly in h, then the Fourier-Bohr spectrum of g is empty.*

Actually we can do a little bit more.

Lemma 1.22. *Let g be a q-multiplicative function bounded by 1. Then the following statements are equivalent.*

1. *The Fourier-Bohr spectrum of g is empty, that is,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n<N} g(n)\, \mathrm{e}\,(-n\beta) = 0.$$

as $\lambda \to \infty$.

2. *We have*

$$\lim_{N\to\infty} \sup_{\beta\in\mathbb{R}} \left| \frac{1}{N} \sum_{n<N} g(n)\, \mathrm{e}\,(-n\beta) \right| = 0.$$

3. *We have*

$$\lim_{\lambda\to\infty} \sup_{h\in\mathbb{Z}} |G_\lambda(h)| = 0.$$

Proof. The implication $3 \to 2$ was proved in the previous section. The converse is trivial, and so is the implication $2 \to 1$. It remains to prove that 1 implies 2.

We prove the special case

$$\sup_{\beta\in\mathbb{R}} \left| \frac{1}{q^\lambda} \sum_{u<q^\lambda} g(u)\, \mathrm{e}\,(-u\beta) \right| \to 0$$

first. For each λ the function

$$\beta \mapsto \left| \frac{1}{q^\lambda} \sum_{u<q^\lambda} g(u)\, \mathrm{e}\,(-u\beta) \right|$$

is continuous and 1-periodic, moreover

$$\lambda \mapsto \left| \frac{1}{q^\lambda} \sum_{u<q^\lambda} g(u)\, \mathrm{e}\,(-u\beta) \right| = \prod_{k<\lambda} \left| \frac{1}{q} \sum_{b<q} g^{(k)}(b)\, \mathrm{e}\left(bq^k\beta\right) \right|$$

is nonincreasing and converges to zero as λ goes to infinity. By Dini's Theorem the convergence is uniform in β. To obtain the full statement, we use (1.17) (with the function $n \mapsto g(n)\, \mathrm{e}\,(-n\beta)$ as g) , (1.18) and Lemma 1.19 again:

$$\sup_{b\in\mathbb{R}} \left| \frac{1}{N} \sum_{n<N} g(n)\, \mathrm{e}\,(-n\beta) \right|^2 \le \frac{q^{2\lambda}}{N^2} + \frac{2q^\lambda}{N} + \frac{2R}{N} + \frac{2R}{q^\lambda}$$

$$+ 2 \sup_{\beta\in\mathbb{R}} \left| q^{-\lambda} \sum_{u<q^\lambda} g(u)\, \mathrm{e}\,(-u\beta)\, \mathrm{e}\left(-huq^{-\lambda}\right) \right|^2 \sum_{h<q^\lambda} \left| \frac{1}{R} \sum_{r<R} \mathrm{e}\left(hrq^{-\lambda}\right) \right|^2 \le \frac{q^{2\lambda}}{N^2} + \frac{2q^\lambda}{N} + \frac{2R}{N} + \frac{2R}{q^\lambda}$$

$$+ 2\frac{q^\lambda + R - 1}{R} \sup_{\beta\in\mathbb{R}} \left| \frac{1}{q^\lambda} \sum_{u<q^\lambda} g(u) e(-u\beta) \right|^2$$

As in the proof of Theorem 1.11 we have to choose R and λ according to the behaviour of the supremum. This finishes the proof. □

So far we have not been concerned with pseudorandomness of an arithmetic function g. As a first statement concerning pseudorandom sequences, we note that pseudorandomness always implies that the spectrum is empty (see Coquet and Mendès France [12]).

Lemma 1.23. *Let g be bounded arithmetic function. If g is pseudorandom (in the sense of Bertrandias), then the Fourier-Bohr spectrum of g is empty.*

Proof. The proof is an application of van der Corput's inequality (Lemma 1.15). We have for all $R \in \{1, \dots, N\}$

$$\left|\frac{1}{N}\sum_{n<N} g(n)\,\mathrm{e}\,(n\beta)\right|^2 \le \frac{N+R}{RN^2}\sum_{|r|<R}\left(1-\frac{|r|}{R}\right)\mathrm{e}\,(r\beta)\sum_{n,n+r<N} g(n+r)\overline{g(n)}$$
$$\ll \frac{1}{R}\sum_{0\le r<R}\left|\frac{1}{N}\sum_{n<N} g(n+r)\overline{g(n)}\right| + O\left(\frac{R}{N}\right).$$

For brevity we write

$$\gamma_r = \lim_{N\to\infty}\frac{1}{N}\sum_{n<N} g(n+r)\overline{g(n)}.$$

Let $\varepsilon \in (0,1)$. By hypothesis we may choose R so large that

$$\frac{1}{R}\sum_{r<R}|\gamma_r| < \varepsilon^2.$$

Moreover, we choose N_0 in such a way that $R/N_0 < \varepsilon^2$ and

$$\left|\frac{1}{N}\sum_{n<N} g(n+r)\overline{g(n)} - \gamma_r\right| < \varepsilon^2$$

for all $r < R$ and $N \ge N_0$. Then for $N \ge N_0$ we have

$$\left|\frac{1}{N}\sum_{n<N} g(n)\,\mathrm{e}\,(n\beta)\right|^2 \ll \frac{1}{R}\sum_{r<R}|\gamma_r| + \frac{1}{R}\sum_{r<R}\left|\frac{1}{N}\sum_{n<N} g(n+r)\overline{g(n)} - \gamma_r\right| + O(R/N_0) < 3\varepsilon^2.$$

□

Remark. We note that the statement of the lemma is similar to van der Corput's theorem in the theory of uniform distribution, which states the following: Let x be a sequence of real numbers such that for all $q \ge 1$ the sequence $(x_{n+q} - x_n)_n$ is uniformly distributed modulo 1. Then x is uniformly distributed modulo 1. (The proof of this theorem is also an application of van der Corput's inequality.) By Weyl's criterion we can rewrite this theorem as follows, writing $g(n) = \mathrm{e}\,(x_n)$: If for all integers $r, h \ge 1$ we have

$$\frac{1}{N}\sum_{n<N} g(n+r)^h\overline{g(n)}^h = o(1),$$

then for all $h \geq 1$

$$\frac{1}{N}\sum_{n<N} g(n)^h = o(1).$$

We refer to Bertrandias [6] for a closer investigation of the connection between pseudorandom sequences and uniform distribution modulo one.

The converse of Lemma 1.23 does not always hold. However, it is true for q-multiplicative functions $g : \mathbb{N} \to \mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$, which has been proved by Coquet (see his thesis [10, p. 23], and [9]). We first establish the existence of the correlation of q-multiplicative functions.

Lemma 1.24. *Let g be a q-multiplicative function $g : \mathbb{N} \to \mathbb{T}$. Then for every $r \geq 0$ the limit*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n<N} g(n+r)\overline{g(n)}$$

exists.

Proof. Let $r, \lambda, N \geq 0$ and set $k = \max\{j : jq^\lambda \leq N\}$. Then by Lemma 1.17 we get

$$\sum_{n<N} g(n+r)\overline{g(n)} = \sum_{n<N} g_\lambda(n+r)\overline{g_\lambda(n)} + O(Nrq^{-\lambda}) = \sum_{n<kq^\lambda} g_\lambda(n+r)\overline{g_\lambda(n)} + O(q^\lambda + Nrq^{-\lambda}).$$

By the q^λ-periodicity of $g_\lambda(n+r)\overline{g_\lambda(n)}$ and since $\left|k/N - q^{-\lambda}\right| \leq 1/N$ we get

$$\left|\frac{1}{N}\sum_{n<N} g(n+r)\overline{g(n)} - \frac{1}{q^\lambda}\sum_{n<q^\lambda} g_\lambda(n+r)\overline{g_\lambda(n)}\right| \ll \frac{q^\lambda}{N} + \frac{r}{q^\lambda}.$$

By the triangle inequality it follows that the values $\frac{1}{N}\sum_{n<N} g(n+r)\overline{g(n)}$ form a Cauchy sequence and therefore a convergent sequence, which proves the existence of the correlation of g. □

The following theorem is due to Coquet [10] and clarifies the connection between the Fourier-Bohr spectrum and pseudorandomness for a class of q-multiplicative functions.

Theorem 1.25 (Coquet)**.** *Let f be a q-additive function and $g(n) = \mathrm{e}(f(n))$. The following statements are equivalent.*

1. *g has empty Fourier-Bohr spectrum.*
2. *g is pseudorandom in the sense of Bertrandias.*
3. *For all real α we have*
$$\sum_{r\geq 0}\sum_{a<q} \|f(aq^r) + \alpha aq^r\|^2 = \infty.$$
4. *We have*
$$\sum_{r\geq 0}\sum_{2\leq a\leq q} \|f(aq^r) - af(q^r)\|^2 = \infty.$$

Using discrete Fourier terms, we want to reprove the implication 1→2. This proof is simpler than the original one [10], which uses the route $1 \to 3 \to 4 \to 2$, moreover it is valid for all bounded q-multiplicative functions g, not only for q-multiplicative functions to $\mathbb{T}$. Since it is our aim to demonstrate the usefulness of Fourier terms for q-multiplicative functions, we will only be concerned with the first two items of this theorem.

Proposition 1.26. *Let g be a bounded q-multiplicative function such that the Fourier coefficients $G_\lambda(h)$ satisfy the estimate*

$$\sup_{h\in\mathbb{Z}} |G_\lambda(h)| = \sup_{h\in\mathbb{Z}} \left| \frac{1}{q^\lambda} \sum_{u<q^\lambda} g(u)\, \mathrm{e}\left(huq^{-\lambda}\right) \right| = o(1).$$

Then g is pseudorandom in the sense of Bertrandias. In particular, if the spectrum of g is empty, this conclusion holds. Moreover, if

$$\sup_h |G_\lambda(h)| \ll \mathrm{e}^{-\eta\lambda}$$

for some $\eta > 0$, then the correlation γ satisfies

$$\frac{1}{R} \sum_{r<R} |\gamma_r| \ll R^{-\varepsilon}$$

for some $\varepsilon > 0$.

Proof. The "in particular" statement follows from Lemma 1.22. Assume that λ is a nonnegative integer. We set

$$g_\lambda(n) = g\left(n \bmod q^\lambda\right).$$

Let $R, k, \lambda \geq 1$. For all $r < R$ choose ε_r on the unit circle in such a way that

$$\varepsilon_r \sum_{0\leq n<kq^\lambda} g_\lambda(n+r)\overline{g_\lambda(n)}$$

is a nonnegative real number. Using Lemma 1.16 and the inequality of Cauchy-Schwarz we get the following bound that is uniform in k.

$$\left(\frac{1}{R} \sum_{0\leq r<R} \left| \frac{1}{kq^\lambda} \sum_{0\leq n<kq^\lambda} g_\lambda(n+r)\overline{g_\lambda(n)} \right| \right)^2$$

$$
\begin{aligned}
&= \left(\frac{1}{R}\sum_{0\le r<R}\varepsilon_r\sum_{0\le h<q^\lambda}|G_\lambda(h)|^2\,\mathrm{e}\left(hrq^{-\lambda}\right)\right)^2\\
&= \frac{1}{R^2}\left|\sum_{0\le h<q^\lambda}|G_\lambda(h)|^2\sum_{0\le r<R}\varepsilon_r e\left(hrq^{-\lambda}\right)\right|^2\\
&\le \frac{1}{R^2}\sum_{0\le h<q^\lambda}|G_\lambda(h)|^4\sum_{0\le h<q^\lambda}\left|\sum_{0\le r<R}\varepsilon_r\,\mathrm{e}\left(hrq^{-\lambda}\right)\right|^2\\
&= \frac{1}{R^2}\sum_{0\le h<q^\lambda}|G_\lambda(h)|^4\sum_{0\le h<q^\lambda}\sum_{0\le r_1,r_2<R}\varepsilon_{r_1}\overline{\varepsilon_{r_2}}\,\mathrm{e}\left(h(r_1-r_2)q^{-\lambda}\right)\\
&= \frac{q^\lambda}{R^2}\sum_{0\le h<q^\lambda}|G_\lambda(h)|^4\sum_{0\le r_1,r_2<R}\varepsilon_{r_1}\overline{\varepsilon_{r_2}}\delta_{r_1,r_2}\\
&= \frac{q^\lambda}{R}\sum_{0\le h<q^\lambda}|G_\lambda(h)|^4.
\end{aligned}
$$

We replace the function g by g_λ, using Lemma 1.17. Therefore we get by taking the limit in k

$$
\begin{aligned}
\frac{1}{R}\sum_{0\le r<R}|\gamma_r| &= \lim_{k\to\infty}\frac{1}{R}\sum_{0\le r<R}\left|\frac{1}{kq^\lambda}\sum_{0\le n<kq^\lambda}g(n+r)\overline{g(n)}\right|\\
&= \lim_{k\to\infty}\frac{1}{R}\sum_{0\le r<R}\left|\frac{1}{kq^\lambda}\sum_{0\le n<kq^\lambda}g_\lambda(n+r)\overline{g_\lambda(n)}\right| + O\left(\frac{R}{q^\lambda}\right) \qquad (1.20)\\
&\le \left(\sum_{0\le h<q^\lambda}|G_\lambda(h)|^4\right)^{1/2}\left(\frac{q^\lambda}{R}\right)^{1/2} + O\left(\frac{R}{q^\lambda}\right),
\end{aligned}
$$

which is valid for $\lambda \ge 0$ and $R \ge 1$. To complete the proof, we make use of Parseval's equality, stating

$$\sum_{h<q^\lambda}|G_\lambda(h)|^2 = 1,$$

and of the uniform convergence to 0 of the Fourier terms. The remaining details are similarly as in the proof of Theorem 1.11 and Lemma 1.22: Assume that $\varepsilon \in (0,1)$ and choose λ_0 so large that $q^{-\lambda} < \varepsilon$ and

$$\left|\sum_{h<q^\lambda}|G_\lambda(h)|^4\right| < \varepsilon^3$$

for all $\lambda \ge \lambda_0$. We choose $R_0 = q^{\lambda_0}$. For $R \ge R_0$ choose $\lambda \ge \lambda_0$ in such a way that

$$\frac{1}{q}\frac{R}{q^\lambda} \le \varepsilon < q\frac{R}{q^\lambda}.$$

We obtain for all $R \geq R_0$

$$\frac{1}{R}\sum_{r<R}|\gamma_r| \leq \varepsilon^{3/2}(q/\varepsilon)^{1/2} + q\varepsilon \leq 2q\varepsilon.$$

In a similar way, the quantitative statement follows from (1.20), noting that we can take a positive power of ε as R_0. □

Proposition 1.26 allows us to prove the following result, which was already proved by Bésineau [7], generalizing earlier results by Mendès France [44].

Corollary 1.27 (Bésineau)**.** *Assume that $q \geq 2$ and let ϑ be a real number such that $(q-1)\vartheta \notin \mathbb{Z}$. Then the function*

$$n \mapsto \mathrm{e}\left(\vartheta s_q(n)\right)$$

is pseudorandom.

Proof. By Lemma 1.12 we get in particular the fact that the Fourier terms $G_\lambda(h)$ for q-multiplicative functions g of the form $n \mapsto \mathrm{e}\left(\vartheta s_q(n)\right)$, where $(q-1)\vartheta \notin \mathbb{Z}$, converge to 0 uniformly in h. Combining this with Proposition 1.26 we see that these functions are pseudorandom. □

To finish this section, we note that pointwise convergence to zero of the Fourier coefficients is not sufficient for pseudorandomness. To see this, consider the q-additive function $n \mapsto n\vartheta$, where $\vartheta \in \mathbb{R} \setminus \bigcup_{k\geq 0} q^{-k}\mathbb{Z}$. Then for each integer h we have

$$\frac{1}{q^\lambda}\sum_{u<q^\lambda}\mathrm{e}\left(u\vartheta - huq^{-\lambda}\right) = o(1)$$

by Lemma 1.18, but ϑ is contained in the spectrum, which is therefore not empty.

1.2.4 The Zeckendorf sum-of-digits function

In this section we study a digital representation different from the ordinary sum-of-digits function in base q, the *Zeckendorf sum-of-digits* function. It is based on a theorem of Zeckendorf [57], stating that every positive integer n can be written in a unique way as the sum of non-adjacent Fibonacci numbers. That is, there are uniquely determined $\varepsilon_i \in \{0,1\}$ such that

$$n = \sum_{n\geq 2}\varepsilon_n F_n \quad \text{and} \quad \left(\varepsilon_n = 1 \Rightarrow \varepsilon_{n+1} = 0\right), \tag{1.21}$$

where $F_0 = 0$, $F_1 = 1$ and $F_{n+2} = F_n + F_{n+1}$ for $n \geq 0$. Using this representation, we define

$$Z(n) = \sum_{n\geq 2}\varepsilon_n,$$

which is the *Zeckendorf sum of digits* of n.

In the previous section (Corollary 1.2.3) we stated a result by Bésineau concerning the pseudorandomness of the functions $n \mapsto \mathrm{e}\left(\vartheta s_q(n)\right)$. We want to show that it is possible to

obtain an analogous result for the Zeckendorf sum-of-digits function, by adapting the method of the previous section, in particular by using discrete Fourier analysis. The purpose of this section is to prove the following theorem.

Theorem 1.28. *Let $\vartheta \in \mathbb{R} \setminus \mathbb{Z}$. Then the function $n \mapsto \mathrm{e}(\vartheta Z(n))$ is pseudorandom.*

The proof is divided into several steps. First, in analogy to the case of the ordinary sum-of-digits function, we define the truncated version Z_λ of Z as follows: If $n = \sum_{k\geq 2} \varepsilon_k F_k$ is the Zeckendorf representation of n, then

$$Z_\lambda = \sum_{2\leq k<\lambda} \varepsilon_k.$$

Similarly to the q-adic case (see Lemma 1.17) the following lemma holds.

Lemma 1.29. *Let $\lambda \geq 2$ be an integer and $N, r \geq 0$. Then*

$$|\{n < N : Z(n+r) - Z(n) \neq Z_\lambda(n+r) - Z_\lambda(n)\}| \leq N\frac{r}{F_{\lambda-1}}.$$

Proof. The proof is along the lines of Lemma 1.17. Let $(w_i)_i$ be the enumeration of the integers n such that $\varepsilon_2(n) = \cdots = \varepsilon_{\lambda-1} = 0$, in increasing order. Then the intervals $[w_i, w_{i+1})$ constitute a partition of the set $\mathbb{N}$ into intervals of length F_λ and $F_{\lambda-1}$, which can be seen as follows. If $\varepsilon_\lambda(w_i) = 1$, then by the monotonicity of the Zeckendorf numeration (taking the lexicographical order) we have $\varepsilon_j(n) = \varepsilon_j(w_i)$ for $w_i \leq n < w_i + F_{\lambda-1}$, since $n - w_i$ only uses Fibonacci numbers up to $F_{\lambda-2}$. The numbers w_i and $n - w_i$ can therefore be added without a "carry", that is, no pair of adjacent Fibonacci numbers appears. In the addition $w_i + F_{\lambda-1}$ however a carry appears and $\varepsilon_i(w_i + F_{\lambda-1}) = 0$ for $i < \lambda$. It follows that $w_{i+1} = w_i + F_{\lambda-1}$. The case that $\varepsilon_\lambda(w_i) = 0$ is similar.

Moreover, for $w_i \leq n < w_{i+1} - r$ we have $\varepsilon_j(n+r) = \varepsilon_j(n)$ for $j \geq \lambda$. It follows that

$$|\{n \in \{w_i, \dots, w_{i+1} - 1\} : Z(n+r) - Z(n) \neq Z_\lambda(n+r) - Z_\lambda(n)\}| \leq r.$$

By concatenating blocks, the statement follows therefore for the case that $N = w_i$ for some i. It remains to treat the case that $w_i < N < w_{i+1}$ for some i. To this end we just note that

$$\frac{1}{N - w_i} |\{w_i, \dots, N-1\} \cap \{w_{i+1} - r, \dots, w_{i+1} - 1\}|$$

decreases for $w_i \leq N < w_{i+1} - r$ and increases for $w_{i+1} - r \leq N < w_{i+1}$. By concatenating blocks, the full statement follows. □

As above, this lemma implies the existence of the autocorrelation of the functions $n \mapsto e(\vartheta Z(n))$. (Note: we do not introduce a direct generalization of the term "q-multiplicative function". Of course an analogous version of Lemma 1.29 for this kind of functions holds.)

Lemma 1.30. *For each $r \geq 0$ and $\vartheta \in \mathbb{R}$ the limit*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n<N} \mathrm{e}\left(\vartheta\left(Z(n+r) - Z(n)\right)\right)$$

exists.

We skip the proof, which is analogous to the proof of Lemma 1.24 and uses Lemma 1.29 instead of Lemma 1.17. Next we prove a uniform upper bound for Fourier like terms $\tilde{G}_k$.

Lemma 1.31. *Let $\varphi = (\sqrt{5}+1)/2$. For $k \geq 0$ we define*

$$\tilde{G}_k = \tilde{G}_k(\vartheta, \beta) = \frac{1}{\varphi^k} \sum_{0 \leq u < F_k} \mathrm{e}\left(\vartheta Z(u) + \beta u\right).$$

Then for $\vartheta \in \mathbb{R} \setminus \mathbb{Z}$ there exists $\eta > 0$ and a constant C such that for all $k \geq 2$ and all $\beta \in \mathbb{R}$ we have

$$\left|\tilde{G}_k(\vartheta, \beta)\right| \leq C\, \mathrm{e}^{-k\eta}.$$

Proof. By the relation $Z(u + F_k) = 1 + Z(u)$ that holds for $k \geq 2$ and $0 \leq u < F_k$ we get for $k \geq 2$

$$\begin{aligned}\tilde{G}_{k+1} = \frac{1}{\varphi^{k+1}} \sum_{u < F_k} \mathrm{e}\left(\vartheta Z(u) + \beta u\right) \\ + \frac{1}{\varphi^{k+1}} \sum_{u < F_{k-1}} \mathrm{e}\left(\vartheta Z(u + F_k) + \beta u + \beta F_k\right) = \frac{1}{\varphi} \tilde{G}_k + \frac{1}{\varphi^2} \mathrm{e}\left(\vartheta + \beta F_k\right) \tilde{G}_{k-1}.\end{aligned}$$

We write $z_k = z_k(\vartheta, \beta) = \mathrm{e}\left(\vartheta + \beta F_k\right)$ and $A_k = A_k(\vartheta, \beta) = \left(\begin{smallmatrix} \varphi^{-1} & \varphi^{-2} z_k \\ 1 & 0 \end{smallmatrix}\right)$. We obtain

$$\begin{pmatrix} \tilde{G}_{k+1} \\ \tilde{G}_k \end{pmatrix} = A_k \begin{pmatrix} \tilde{G}_k \\ \tilde{G}_{k-1} \end{pmatrix}.$$

A short calculation reveals that

$$\begin{pmatrix} \tilde{G}_{k+5} \\ \tilde{G}_{k+4} \end{pmatrix} = A_{k+4} A_{k+3} A_{k+2} A_{k+1} A_k \begin{pmatrix} \tilde{G}_k \\ \tilde{G}_{k-1} \end{pmatrix} = \begin{pmatrix} a_k & b_k \\ c_k & d_k \end{pmatrix} \begin{pmatrix} \tilde{G}_k \\ \tilde{G}_{k-1} \end{pmatrix}$$

for $k \geq 2$, where

$$\begin{aligned} a_k &= \varphi^{-5} \left(1 + z_{k+1} + z_{k+2} + z_{k+3}(1 + z_{k+1}) + z_{k+4}(1 + z_{k+1} + z_{k+2})\right), \\ b_k &= \varphi^{-6} z_k \left(1 + z_{k+2} + z_{k+3} + z_{k+4}(1 + z_{k+2})\right), \\ c_k &= \varphi^{-4} \left(1 + z_{k+1} + z_{k+2} + z_{k+3}(1 + z_{k+1})\right), \\ d_k &= \varphi^{-5} z_k \left(1 + z_{k+2} + z_{k+3}\right). \end{aligned}$$

To obtain the result, we use the row-sum norm $\|\cdot\|_\infty$ for matrices, which is derived from the maximum norm for vectors and which is sub-multiplicative. Since $\|A_k\|_\infty \leq 1$ it suffices to prove that

$$\sup_{\substack{\beta \in \mathbb{R} \\ k \geq 2}} \|A_{k+4} A_{k+3} A_{k+2} A_{k+1} A_k\|_\infty < 1. \tag{1.22}$$

By Lemma 1.4 and the elementary estimate $\cos(2\pi x) \leq 1 - 2\|x\|^2$ we get

$$k - |1 + \mathrm{e}(x_1) + \cdots + \mathrm{e}(x_{k-1})| \geq \frac{1}{2} \max_{1 \leq i < k} (1 - \Re\, \mathrm{e}(x_i)) \geq \max_{1 \leq i < k} \|x_i\|^2 \tag{1.23}$$

for any family $(x_1, \dots, x_{k-1})$ of real numbers. Note also that $|a_k| \le 8\varphi^{-5}$, $|b_k| \le 5\varphi^{-6}$, $|c_k| \le 5\varphi^{-4}$ and $|d_k| \le 3\varphi^{-5}$. We assume that $|a_k| + |b_k| \ge 1 - \varepsilon$. Then $8\varphi^{-5} - |a_k| \le \varepsilon$ and $5\varphi^{-6} - |b_k| \le \varepsilon$, since in the other case we get $|a_k| + |b_k| < 8\varphi^{-5} + 5\varphi^{-6} - \varepsilon = 1 - \varepsilon$, which contradicts our assumption. By (1.23) it follows therefore that

$$\max\{\|\vartheta + \beta F_{k+1}\|^2, \|\vartheta + \beta F_{k+2}\|^2, \|\vartheta + \beta F_{k+3}\|^2\} \le 8 - \varphi^5 |a_k| \le \varphi^5 \varepsilon.$$

By an analogous argument and $\varphi \ge 1$, the assumption $|c_k| + |d_k| \ge 1 - \varepsilon$ leads to the same estimate. Now if $\|A_{k+4}A_{k+3}A_{k+2}A_{k+1}A_k\|_\infty \ge 1 - \varepsilon$, we have $|a_k| + |b_k| \ge 1 - \varepsilon$ or $|c_k| + |d_k| \ge 1 - \varepsilon$ and therefore

$$\begin{aligned}\|\vartheta\| &= \|\vartheta + \beta F_{k+1} + \vartheta + \beta F_{k+2} - (\vartheta + \beta F_{k+3})\| \\ &\le \|\vartheta + \beta F_{k+1}\| + \|\vartheta + \beta F_{k+2}\| + \|\vartheta + \beta F_{k+3}\| \le 3\sqrt{\varphi^5 \varepsilon}.\end{aligned}$$

If we assume therefore that (1.22) is violated, we get $\vartheta \in \mathbb{Z}$, a case that is excluded. □

Now we are ready to present the proof of Theorem 1.28.

Let $\lambda \ge 2$ and assume that $(w_i)_i$ be the sequence from the proof of 1.29, that is, w_i enumerates the integers n such that $\varepsilon_2(n) = \cdots = \varepsilon_{\lambda-1}(n) = 0$ in increasing order.

We define

$$G_\lambda(h) = \frac{1}{F_\lambda} \sum_{u < F_\lambda} \mathrm{e}\left(\vartheta Z_\lambda(u) - \frac{hu}{F_\lambda}\right).$$

By Lemma 1.31 and since $F_\lambda \asymp \varphi^\lambda$, we obtain the bound

$$|G_\lambda(h)| \le C e^{-\eta\lambda} \tag{1.24}$$

for some $\eta > 0$ that is uniform in h.

Let $n < F_\lambda$. Then by inverting the Fourier transform we have

$$\mathrm{e}\left(\vartheta Z_\lambda(n)\right) = \sum_{h < F_\lambda} G_\lambda(h)\, \mathrm{e}\left(hnF_\lambda^{-1}\right)$$

and

$$\mathrm{e}\left(\vartheta Z_\lambda(-n)\right) = \sum_{h < F_\lambda} \overline{G_\lambda(-h)}\, \mathrm{e}\left(hnF_\lambda^{-1}\right).$$

Let i be such that $w_{i+1} - w_i = F_\lambda$ and assume that $r \ge 0$. We have

$$\begin{aligned}\sum_{h<F_\lambda} \mathrm{e}\left(\frac{hr}{F_\lambda}\right) |G_\lambda(h)|^2 &= \frac{1}{F_\lambda} \sum_{u,v<F_\lambda} \mathrm{e}\left(\vartheta Z_\lambda(u) - \vartheta Z_\lambda(v)\right) \frac{1}{F_\lambda} \sum_{h<F_\lambda} \mathrm{e}\left(\frac{h}{F_\lambda}(v + r - u)\right) \\ &= \frac{1}{F_\lambda} \sum_{u,v<F_\lambda} [v + r \equiv u \bmod F_\lambda]\, \mathrm{e}\left(\vartheta(Z_\lambda(u) - Z_\lambda(v))\right) \\ &= \frac{1}{F_\lambda} \sum_{w_i \le u,v < w_{i+1}} [v + r \equiv u \bmod F_\lambda]\, \mathrm{e}\left(\vartheta(Z_\lambda(u) - Z_\lambda(v))\right) \\ &= \frac{1}{F_\lambda} \sum_{w_i \le u < w_{i+1} - r} \mathrm{e}\left(\vartheta(Z_\lambda(v + r) - Z_\lambda(v))\right) + O\left(\frac{r}{F_\lambda}\right) \\ &= \frac{1}{F_\lambda} \sum_{w_i \le u < w_{i+1}} \mathrm{e}\left(\vartheta(Z_\lambda(v + r) - Z_\lambda(v))\right) + O\left(\frac{r}{F_\lambda}\right).\end{aligned}$$

We write $g_\lambda(n) = \mathrm{e}(\vartheta Z_\lambda(n))$. Let $\ell \geq 0$. We denote by a the number of $i < \ell$ such that $w_{i+1} - w_i = F_{\lambda-1}$ and by b the number of $i < \ell$ such that $w_{i+1} - w_i = F_\lambda$.

Since the left hand side (of the calculation above) and the error term are independent of i, we can calculate

$$\begin{aligned}\frac{1}{R}\sum_{0\le r<R}\left|\frac{1}{w_\ell}\sum_{0\le n<w_\ell} g_\lambda(n+r)\overline{g_\lambda(n)}\right| &= \left|\frac{aF_{\lambda-1}}{w_\ell}\frac{1}{R}\sum_{0\le r<R}\varepsilon_r\sum_{0\le h<F_{\lambda-1}}|G_{\lambda-1}(h)|^2\,\mathrm{e}\left(\frac{hr}{F_{\lambda-1}}\right)\right.\\ &\quad\left.+\frac{bF_\lambda}{w_\ell}\frac{1}{R}\sum_{0\le r<R}\varepsilon_r\sum_{0\le h<F_\lambda}|G_\lambda(h)|^2\,\mathrm{e}\left(\frac{hr}{F_\lambda}\right)\right| + O\left(\frac{ar}{w_\ell}+\frac{br}{w_\ell}\right)\\ &= O\left(\frac{r}{F_\lambda}\right)+\frac{1}{R}\left|\frac{aF_{\lambda-1}}{w_\ell}\sum_{0\le h<F_{\lambda-1}}|G_{\lambda-1}(h)|^2\sum_{0\le r<R}\varepsilon_r e\left(\frac{hr}{F_{\lambda-1}}\right)\right.\\ &\quad\left.+\frac{bF_\lambda}{w_\ell}\sum_{0\le h<F_\lambda}|G_\lambda(h)|^2\sum_{0\le r<R}\varepsilon_r e\left(\frac{hr}{F_\lambda}\right)\right| \le O\left(\frac{r}{F_\lambda}\right)\\ &\quad+\frac{1}{R}\left|\sum_{0\le h<F_{\lambda-1}}|G_{\lambda-1}(h)|^2\sum_{0\le r<R}\varepsilon_r e\left(\frac{hr}{F_{\lambda-1}}\right)\right|+\frac{1}{R}\left|\sum_{0\le h<F_\lambda}|G_\lambda(h)|^2\sum_{0\le r<R}\varepsilon_r e\left(\frac{hr}{F_\lambda}\right)\right|.\end{aligned}$$

By Cauchy-Schwarz we obtain

$$\begin{aligned}&\frac{1}{R^2}\left|\sum_{0\le h<F_\lambda}|G_\lambda(h)|^2\sum_{0\le r<R}\varepsilon_r e\left(\frac{hr}{F_\lambda}\right)\right|^2\\ &\qquad\le \frac{1}{R^2}\sum_{0\le h<F_\lambda}|G_\lambda(h)|^4\sum_{0\le h<F_\lambda}\left|\sum_{0\le r<R}\varepsilon_r\,\mathrm{e}\left(\frac{hr}{F_\lambda}\right)\right|^2\\ &\qquad\le \frac{1}{R^2}\sum_{0\le h<F_\lambda}|G_\lambda(h)|^4\sum_{0\le h<F_\lambda}\sum_{0\le r_1,r_2<R}\varepsilon_{r_1}\overline{\varepsilon_{r_2}}\,\mathrm{e}\left(h\frac{r_1-r_2}{F_\lambda}\right)\\ &\qquad= \frac{F_\lambda}{R^2}\sum_{0\le h<F_\lambda}|G_\lambda(h)|^4\sum_{0\le r_1,r_2<R}\varepsilon_{r_1}\overline{\varepsilon_{r_2}}\delta_{r_1,r_2}\\ &\qquad= \frac{F_\lambda}{R}\sum_{0\le h<F_\lambda}|G_\lambda(h)|^4,\end{aligned}$$

similarly for $\lambda - 1$.

Using Lemma 1.29 and writing $g(n) = \mathrm{e}(\vartheta Z(n))$, we get

$$\begin{aligned}\frac{1}{R}\sum_{0\le r<R}|\gamma_r| &= \lim_{\ell\to\infty}\frac{1}{R}\sum_{0\le r<R}\left|\frac{1}{w_\ell}\sum_{0\le n<w_\ell} g(n+r)\overline{g(n)}\right| \\ &= O\left(\frac{R}{F_\lambda}\right) + \lim_{k\to\infty}\frac{1}{R}\sum_{0\le r<R}\left|\frac{1}{w_\ell}\sum_{0\le n<w_\ell} g_\lambda(n+r)\overline{g_\lambda(n)}\right| \le O\left(\frac{R}{F_\lambda}\right) \\ &+ \left(\left(\sum_{0\le h<F_{\lambda-1}}|G_{\lambda-1}(h)|^4\right)^{1/2} + \left(\sum_{0\le h<F_\lambda}|G_\lambda(h)|^4\right)^{1/2}\right)\left(\frac{F_\lambda}{R}\right)^{1/2}.\end{aligned}$$

By (1.24) we have $\sup_h |G_\lambda(h)| \le Ce^{-\eta\lambda}$ for some $\eta > 0$. Furthermore, since

$$\sum_{h<F_\lambda}|G_\lambda(h)|^2 = 1$$

we obtain

$$\sum_{h<F_\lambda}|G_\lambda(h)|^4 \le C^2e^{-2\eta\lambda}.$$

In the same way as in the proof of Proposition 1.26 we conclude that

$$\frac{1}{R}\sum_{0\le r<R}|\gamma_r| = o(1)$$

as $R\to\infty$. This completes the proof of Theorem 1.28.

1.3 Different bases

1.3.1 Statistical independence of different bases

In section 1.2 we were concerned with the investigation of properties of a single q-multiplicative function g, showing the usefulness of Fourier terms in this context. In this short section we want to combine q_i-multiplicative functions with respect to different bases. This area of research was initiated by the article [29] of Gelfond, in which the following question was posed:

Let $q_1, q_2 \ge 2$, $m_1, m_2 \ge 1$ and l_1, l_2 be integers such that $(q_1, q_2) = 1$, $(m_1, q_1 - 1) = 1$ and $(m_2, q_2 - 1) = 1$. Prove that there exists an $\varepsilon > 0$ such that

$$|\{n \le x : s_{q_1}(n) \equiv l_1 \bmod m_1 \text{ and } s_{q_2}(n) \equiv l_2 \bmod m_2\}| = \frac{x}{m_1m_2} + O(x^{1-\varepsilon}). \tag{1.25}$$

This question, which can be viewed as a problem on the "independence" of digital representation in coprime bases, was answered by Kim [33], although a non-quantitative version of this result was proved much earlier by Bésineau [7]. Note that this problem is concerned

with special q-multiplicative functions, namely $n \mapsto \mathrm{e}(\vartheta s_q(n))$ for some ϑ. Coquet [10] studied even the more general case of general q-multiplicative functions and proved statements on the independence of q-multiplicative functions with respect to different bases. In fact, Bésineau's result, a non-quantitative version of (1.25), can be proved with the help of the following result from [10].

Theorem 1.32 (Coquet)**.** *Let $q_1, q_2 \geq 2$ be coprime integers and $g_i : \mathbb{N} \to \mathbb{T}$ be q_i-multiplicative. The following statements are equivalent.*

1. *At least one of the functions g_i is pseudorandom*
2. *$g_1 g_2$ is pseudorandom*
3. *$g_1 g_2$ has empty spectrum*
4. *At least one of the functions g_i has empty spectrum.*

We will show that also in this case Fourier terms for g_1 and g_2 are a useful tool. More specifically, we want to give a new proof of the implication $4 \to 2$ using the discrete Fourier transform. In fact, a modification of this proof even allows us to obtain an alternative proof of the *quantitative* result (1.25). The general implication $2 \to 3$ follows from Lemma 1.23 and does not have anything to do with the q_i-multiplicativity of the functions involved. The implication $3 \to 4$ is shown by means of a Turán-Kubilius type inequality (Proposition I.2 in Coquet's thesis [10]; a similar version of this can be found in [21]). We do not reproduce the proof of this implication here. The equivalence $4 \leftrightarrow 1$ follows directly from Theorem 1.25 and Lemma 1.23.

Proof of the implication 4 → 2. It is sufficient to assume that the spectrum of g_1 is empty, which means, by Lemma 1.22, that the Fourier coefficients $G_{1,\lambda_1}(h)$ converge to zero uniformly in h.

First of all, we note that the correlation of $g_1 g_2$ exists, which can be shown in a way that is similar to the case of a single base. More specifically, we can use lemma 1.17 and the Cauchy criterion again.

We follow the calculation from the proof of Proposition 1.26 concerning the case of a single base. By a straightforward calculation we get for all $s \geq 1$

$$\frac{1}{sq_1^{\lambda_1}q_2^{\lambda_2}} \sum_{n<sq_1^{\lambda_1}q_2^{\lambda_2}} g_{1,\lambda_1}(n+r)g_{2,\lambda_2}(n+r)\overline{g_{1,\lambda_1}(n)g_{2,\lambda_2}(n)}$$
$$= \sum_{\substack{h_1<q_1^{\lambda_1}\\ h_2<q_2^{\lambda_2}}} |G_{1,\lambda_1}(h_1)|^2 |G_{2,\lambda_2}(h_2)|^2 \mathrm{e}\left(r\left(\frac{h_1}{q_1^{\lambda_1}}+\frac{h_2}{q_2^{\lambda_2}}\right)\right) \quad (1.26)$$

and therefore we get, choosing ε_r appropriately for $r < R$ and applying Cauchy-Schwarz on the sum over h_1

$$\left(\frac{1}{R}\sum_{r<R}\left|\frac{1}{sq_1^{\lambda_1}q_2^{\lambda_2}} \sum_{n<sq_1^{\lambda_1}q_2^{\lambda_2}} g_{1,\lambda_1}(n+r)g_{2,\lambda_2}(n+r)\overline{g_{1,\lambda_1}(n)g_{2,\lambda_2}(n)}\right|\right)^2 \quad (1.27)$$

$$
\begin{aligned}
&= \frac{1}{R^2} \left| \sum_{\substack{h_1 < q_1^{\lambda_1} \\ h_2 < q_2^{\lambda_2}}} |G_{1,\lambda_1}(h_1)|^2 \, |G_{2,\lambda_2}(h_2)|^2 \sum_{r<R} \varepsilon_r \, \mathrm{e}\left(r \left(\frac{h_1}{q_1^{\lambda_1}} + \frac{h_2}{q_2^{\lambda_2}} \right) \right) \right|^2 \\
&\leq \frac{1}{R^2} \sum_{h < q_1^{\lambda_1}} |G_{1,\lambda_1}(h)|^4 \sum_{h_1 < q_1^{\lambda_1}} \left| \sum_{h_2 < q_2^{\lambda_2}} |G_{2,\lambda_2}(h_2)|^2 \sum_{r<R} \varepsilon_r \, \mathrm{e}\left(r \left(\frac{h_1}{q_1^{\lambda_1}} + \frac{h_1}{q_2^{\lambda_2}} \right) \right) \right|^2 \\
&= \frac{1}{R^2} \sum_{h < q_1^{\lambda_1}} |G_{1,\lambda_1}(h)|^4 \sum_{h_1 < q_1^{\lambda_1}} \sum_{h_2, h_2' < q_2^{\lambda_2}} |G_{2,\lambda_2}(h_2)|^2 \, |G_{2,\lambda_2}(h_2')|^2 \sum_{r,r'<R} \varepsilon_r \overline{\varepsilon_{r'}} \\
&\times \mathrm{e}\left(r \left(\frac{h_1}{q_1^{\lambda_1}} + \frac{h_2}{q_2^{\lambda_2}} \right) \right) \mathrm{e}\left(-r' \left(\frac{h_1}{q_1^{\lambda_1}} + \frac{h_2'}{q_2^{\lambda_2}} \right) \right) \\
&= \frac{1}{R^2} \sum_{h < q_1^{\lambda_1}} |G_{1,\lambda_1}(h)|^4 \sum_{h_2, h_2' < q_2^{\lambda_2}} |G_{2,\lambda_2}(h_2)|^2 \, |G_{2,\lambda_2}(h_2')|^2 \sum_{r,r'<R} \varepsilon_r \overline{\varepsilon_{r'}} \\
&\times \mathrm{e}\left(r \left(\frac{h_2}{q_2^{\lambda_2}} + \frac{h_2'}{q_2^{\lambda_2}} \right) \right) \sum_{h_1 < q_1^{\lambda_1}} \mathrm{e}\left((r - r') \frac{h_1}{q_1^{\lambda_1}} \right) \\
&\leq \frac{q_1^{\lambda_1}}{R} \sum_{h < q_1^{\lambda_1}} |G_{1,\lambda_1}(h)|^4 ,
\end{aligned}
$$

the last step being justified by orthogonality relations and Parseval's identity. We obtain for all $\lambda_1, \lambda_2 \geq 0$, using Lemma 1.17,

$$
\begin{aligned}
&\frac{1}{R} \sum_{0 \leq r < R} \left| \lim_{N \to \infty} \frac{1}{N} \sum_{0 \leq n < N} g_1(n+r) g_2(n+r) \overline{g_1(n) g_2(n)} \right| \\
&\quad = \lim_{s \to \infty} \frac{1}{R} \sum_{0 \leq r < R} \left| \frac{1}{s q_1^{\lambda_1} q_2^{\lambda_2}} \sum_{0 \leq n < s q_1^{\lambda_1} q_2^{\lambda_2}} g_1(n+r) g_2(n+r) \overline{g_1(n) g_2(n)} \right| = O\left(\frac{R}{q_1^{\lambda_1}} + \frac{R}{q_2^{\lambda_2}} \right) \\
&\qquad + \lim_{s \to \infty} \frac{1}{R} \sum_{0 \leq r < R} \left| \frac{1}{s q_1^{\lambda_1} q_2^{\lambda_2}} \sum_{0 \leq n < s q_1^{\lambda_1} q_2^{\lambda_2}} g_{1,\lambda_1}(n+r) g_{2,\lambda_2}(n+r) \overline{g_{1,\lambda_1}(n) g_{2,\lambda_2}(n)} \right| \\
&\qquad\qquad \leq \left(\sum_{h_1 < q_1^{\lambda_1}} |G_{1,\lambda_1}(h_1)|^4 \right)^{1/2} \left(\frac{q_1^{\lambda_1}}{R} \right)^{1/2} + O\left(\frac{R}{q_1^{\lambda_1}} + \frac{R}{q_2^{\lambda_2}} \right).
\end{aligned}
$$

Filling in the missing details in the same way as in the proof of Proposition 1.26 the proof of the implication $1 \to 2$. □

Additionally, we want to show directly that 2 implies 4.

Proof of the implication 2 → 4. The proof is by contradiction. Assume that each of g_1 and g_2 has nonempty spectrum. Then $G_{1,\lambda_1}(h_1)$ and $G_{2,\lambda_2}(h_2)$ do not converge to 0 uniformly, which follows from Lemma 1.22. By (1.26) we get for all $\lambda_1, \lambda_2 \geq 0$

$$\begin{aligned}
&\frac{1}{R}\sum_{r<R}\left|\frac{1}{sq_1^{\lambda_1}q_2^{\lambda_2}}\sum_{n<sq_1^{\lambda_1}q_2^{\lambda_2}} g_{1,\lambda_1}(n+r)g_{2,\lambda_2}(n+r)\overline{g_{1,\lambda_1}(n)g_{2,\lambda_2}(n)}\right|^2\\
\geq&\frac{1}{R}\sum_{|r|<R}\left(1-\frac{|r|}{R}\right)\left|\sum_{\substack{h_1<q_1^{\lambda_1}\\ h_2<q_2^{\lambda_2}}}|G_{1,\lambda_1}(h_1)|^2\,|G_{2,\lambda_2}(h_2)|^2\,\mathrm{e}\left(r\left(\frac{h_1}{q_1^{\lambda_1}}+\frac{h_2}{q_2^{\lambda_2}}\right)\right)\right|^2\\
=&\frac{1}{R}\sum_{\substack{h_1,k_1<q_1^{\lambda_1}\\ h_2,k_2<q_2^{\lambda_2}}}|G_{1,\lambda_1}(h_1)|^2\,|G_{1,\lambda_1}(k_1)|^2\,|G_{2,\lambda_2}(h_2)|^2\,|G_{2,\lambda_2}(k_2)|^2\\
&\times\sum_{|r|<R}\left(1-\frac{|r|}{R}\right)\mathrm{e}\left(r\left(\frac{h_1-k_1}{q_1^{\lambda_1}}+\frac{h_2-k_2}{q_2^{\lambda_2}}\right)\right)\\
=&\sum_{\substack{h_1,k_1<q_1^{\lambda_1}\\ h_2,k_2<q_2^{\lambda_2}}}|G_{1,\lambda_1}(h_1)|^2\,|G_{1,\lambda_1}(k_1)|^2\,|G_{2,\lambda_2}(h_2)|^2\,|G_{2,\lambda_2}(k_2)|^2\\
&\times\left|\frac{1}{R}\sum_{r<R}\mathrm{e}\left(r\left(\frac{h_1-k_1}{q_1^{\lambda_1}}+\frac{h_2-k_2}{q_2^{\lambda_2}}\right)\right)\right|^2\\
\geq&\sum_{\substack{h_1<q_1^{\lambda_1}\\ h_2<q_2^{\lambda_2}}}|G_{1,\lambda_1}(h_1)|^4\,|G_{2,\lambda_2}(h_2)|^4.
\end{aligned}$$

and therefore

$$\begin{aligned}
\frac{1}{R}\sum_{0\leq r<R}\left|\lim_{N\to\infty}\frac{1}{N}\sum_{0\leq n<N}g_1(n+r)g_2(n+r)\overline{g_1(n)g_2(n)}\right| &= O\left(\frac{R}{q_1^{\lambda_1}}+\frac{R}{q_2^{\lambda_2}}\right)\\
+\lim_{s\to\infty}\frac{1}{R}\sum_{0\leq r<R}\left|\frac{1}{sq_1^{\lambda_1}q_2^{\lambda_2}}\sum_{0\leq n<sq_1^{\lambda_1}q_2^{\lambda_2}} g_{1,\lambda_1}(n+r)g_{2,\lambda_2}(n+r)\overline{g_{1,\lambda_1}(n)g_{2,\lambda_2}(n)}\right|&\\
\geq\sum_{\substack{h_1<q_1^{\lambda_1}\\ h_2<q_2^{\lambda_2}}}|G_{1,\lambda_1}(h_1)|^4\,|G_{2,\lambda_2}(h_2)|^4+O\left(\frac{R}{q_1^{\lambda_1}}+\frac{R}{q_2^{\lambda_2}}\right)&
\end{aligned}$$

with an implied constant bounded by 1. Let $\varepsilon > 0$ be such that $\sup_{h_1<q_1^{\lambda_1}}|G_{1,\lambda_1}(h_1)|^4 \geq \varepsilon$ for infinitely many λ_1 and $\sup_{h_2<q_2^{\lambda_2}}|G_{2,\lambda_2}(h_2)|^4 \geq \varepsilon$ for infinitely many λ_2. Let R be given.

Choose λ_1, λ_2 such that the above inequalities are satisfied and in such a way that $R/q_1^{\lambda_1} < \varepsilon^2/3$ and $R/q_1^{\lambda_1} < \varepsilon^2/3$. Then the left hand side is bounded below by $\varepsilon^2/3$. This contradicts the pseudorandomness of $g_1 g_2$, which finishes the proof. □

As we noted above, Theorem 1.32 can be used to prove Bésineau's result. To this end, we only note that we can use Lemma 1.12, the implication $4 \to 3$ and an argument similar to (1.15).

More importantly, we want to give an alternative proof of the full (quantitative) result (1.25). Assume that N is a positive integer and $\lambda_1, \lambda_2 \geq 0$. Let $g_1(n) = \mathrm{e}(\vartheta_1 s_{q_1}(n))$ and $g_2(n) = \mathrm{e}(\vartheta_2 s_{q_2}(n))$ be such that $\vartheta_1(q_1 - 1) \notin \mathbb{Z}$ and $\vartheta_2(q_2 - 1) \notin \mathbb{Z}$. We round off N to the nearest multiple M of $q_1^{\lambda_1} q_2^{\lambda_2}$ in exchange for an error term $O(q_1^{\lambda_1} q_2^{\lambda_2})$ and apply van der Corput's inequality (Lemma 1.15) for $1 \leq R \leq N$, Lemma 1.17 and (1.27):

$$\begin{aligned}\frac{1}{N^2}\left|\sum_{n<M} g_1(n)g_2(n)\right|^2 &\ll \frac{1}{R}\sum_{|r|<R}\left|\frac{1}{N}\sum_{0\leq n,n+r<M} g_1(n+r)g_2(n+r)\overline{g_1(n)g_2(n)}\right| \\ &\ll \frac{1}{R}\sum_{r<R}\left|\frac{1}{N}\sum_{n<M} g_{1,\lambda_1}(n+r)g_{2,\lambda_2}(n+r)\overline{g_{1,\lambda_1}(n)g_{2,\lambda_2}(n)}\right| + O\left(R\left(\frac{1}{N}+\frac{1}{q_1^{\lambda_1}}+\frac{1}{q_2^{\lambda_2}}\right)\right) \\ &\leq \frac{q_1^{\lambda_1}}{R}\sum_{h<q_1^{\lambda_1}} |G_{1,\lambda_1}(h)|^4 + O\left(R\left(\frac{1}{N}+\frac{1}{q_1^{\lambda_1}}+\frac{1}{q_2^{\lambda_2}}\right)\right).\end{aligned}$$

Moreover, we have to replace the summation over $n < M$ by a summation over $n < N$, which contributes an error term $O\left(q_1^{2\lambda_1} q_2^{2\lambda_2}/N^2 + q_1^{\lambda_1} q_2^{\lambda_2}/N\right)$. By an argument as in the proof of Theorem 1.11 and by Lemma 1.12 we conclude that

$$\frac{1}{N}\sum_{n<N} g_1(n)g_2(n) \ll N^{-\eta}$$

for some $\eta > 0$. Transferring this to a statement on distribution in residue classes (as in (1.15), we obtain (1.25).

Chapter 2

Correlations for the sum-of-digits function

In chapter 1 we were concerned with q-multiplicative functions g. In particular, we reproved that the correlations

$$\gamma_t = \lim_{N\to\infty} \frac{1}{N} \sum_{n<N} g_{n+t}\overline{g_n}$$

always exist for these functions (see Lemma 1.24) and (re)studied the relation of the mean value of the correlation function $t \mapsto \gamma_t$ to the Fourier-Bohr spectrum of g (see Theorem 1.25).

In this chapter we want to specialize to a class of q-multiplicative functions coming from the sum-of-digits function s_q. That is, we study the correlation $\gamma_t(\vartheta)$ of the function $n \mapsto \mathrm{e}(\vartheta s_q(n))$. We prove that this correlation satisfies a surprising reflection property, more precisely we prove that

$$\gamma_t(\vartheta) = \gamma_{t^R}(\vartheta), \tag{2.1}$$

where t^R is the *digit reversal* of t in base q.

Moreover, we prove an analogous property for sequences z satisfying $z_{2t} = z_t$ and $z_{2t+1} = \alpha z_t + \beta z_{t+1}$. This result generalizes (2.1) in the case that $q = 2$, since in this case the correlation satisfies a recurrence relation of the above form.

Finally, we present a series of small observations that we encountered during our quest of understanding sequences z satisfying the above recurrence relation.

This chapter emerged from the paper "A reverse order property of correlation measures of the sum-of-digits function" [47], which is joint work with Johannes Morgenbesser and which has been published in the journal "Integers". That article contains a proof of (2.1), which we reproduce in section 2.2.2.

2.1 Introduction and main results

The letter q always denotes an integer greater than or equal to 2. We will use it exclusively to denote the base of a digital representation. As before, let s_q be the sum-of-digits function in base q.

We repeat the definition of the correlation of a complex valued sequence: Let $z = (z_n)_{n\geq 0}$ be a sequence of complex numbers. We say that the *correlation* of z exists if for all $t \in \mathbb{N}$ the limit

$$\gamma_t = \lim_{N\to\infty} \frac{1}{N} \sum_{n<N} z_{n+t}\overline{z_n} \tag{2.2}$$

exists. In this case, we call γ the *correlation* of z. For completeness, we note [10, section 2.3] that the correlation γ, if it exists, admits a representation as a Fourier transform:

$$\gamma_t = \int_{\mathbb{T}} \mathrm{e}\left(t\vartheta\right)\, \mathrm{d}\mu(\vartheta).$$

The measure μ is a finite measure on the torus $\mathbb{T}$, the so-called *spectral measure* of the sequence z_t. In Chapter 1 we have reproved the known result that the correlation of q-multiplicative functions $g : \mathbb{N} \to \mathbb{T}$ always exists. The proof uses Lemma 1.17, which states that $g(n+r)\overline{g(n)}$ can be replaced by the q^λ-periodic function $g_\lambda(n+r)\overline{g_\lambda(n)}$ in most cases (this was done by Coquet [10] in a similar way). We can therefore study the spectral measure of q-multiplicative functions. To this end, we refer to the article [11] by Coquet, Kamae and Mendès France.

We also note that Mahler [37] proved that the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n<N} (-1)^{s_2(n+t)+s_2(n)}$$

exists, moreover he showed that this limit is nonzero for infinitely many t. These limits form the correlation for the Thue-Morse sequence $g(n) = (-1)^{s_2(n)}$. The existence of the correlation in this case can be linked (see [38]) to dynamical properties of the Thue-Morse dynamical system. However, we will not discuss this approach.

In order to formulate the main results, we define the *reflection* of a nonnegative integer t in base q: Assume that $t \geq 1$ and $t = (\varepsilon_\nu \varepsilon_{\nu-1} \dots \varepsilon_0)_q$ is the proper representation of t in base q, that is, $\nu \geq 0$ is chosen minimal such that $t = \sum_{i\leq\nu} \varepsilon_i q^i$ with certain $\varepsilon_i \in \{0, \dots, q-1\}$. Then $\varepsilon_\nu \neq 0$. We set

$$t^R = \sum_{i\leq\nu} \varepsilon_{\nu-i} q^i = (\varepsilon_0 \varepsilon_1 \dots \varepsilon_\nu)_q .$$

The integer t^R is therefore obtained from t by reversing the order of the digits of t in base q. Note that $0^R = 0$. Clearly for $(q,t) = 1$ we have $t^{RR} = t$, more general if $t = q^k \cdot \hat{t}$ with $(\hat{t}, q) = 1$ and $k \geq 0$, then $t^{RR} = \hat{t}$.

The precise formulation of the first main result of this chapter is the following.

Theorem 2.1 (Morgenbesser and Spiegelhofer [47]). *For $q \geq 2$, $\vartheta \in \mathbb{R}$ and $t \geq 0$ set*

$$\gamma_t(\vartheta) = \lim_{N\to\infty} \frac{1}{N} \sum_{n<N} e(\vartheta(s_q(n+t) - s_q(n)).$$

Then we have

$$\gamma_t(\vartheta) = \gamma_{t^R}(\vartheta),$$

where the reflection t^R has to be taken with respect to the base q.

We note that the defining expression for $\gamma_t(\vartheta)$ is the correlation of the q-multiplicative function $n \mapsto \mathrm{e}(\vartheta s_q(n))$, which exists by Lemma 1.24. Using this theorem we will prove that that the sets

$$A_{k,t} = \{n : s_2(n+t) - s_2(n) = k\}$$

and

$$A_{k,t^R} = \{n : s_2(n+t^R) - s_2(n) = k\}$$

have the same asymptotic densities $\delta(k,t)$ resp. $\delta(k,t^R)$ for all $k \in \mathbb{Z}$.

Theorem 2.2. *Let $q \geq 2$, $k \in \mathbb{Z}$ and $t \geq 0$. Then we have $\delta(k,t) = \delta(k,t^R)$.*

This is a curious result, since at first sight the values of $s_q(n+t)$ and $s_q(n+t^R)$ seem to be unreleated to each other. In fact it is not obvious how to construct a (density-preserving) bijection $\sigma : \mathbb{N} \to \mathbb{N}$ in such a way that $s_q(n+t) = s_q(\sigma(n)+t^R)$. The existence of such a bijection is a corollary of the theorem. However, we prove the theorem in a different way. The result can also be restated in terms of divisibility of binomial coefficients or in terms of *carries* via the use of classical statements by Kummer and Legendre.

Lemma 2.3 (Kummer [35])**.** *Let p be a prime number. The maximum*

$$\max\left\{k : p^k \mid \binom{n}{k}\right\}$$

is equal to the number of carries in the addition of $n-k$ and k in base p.

Lemma 2.4 (Legendre)**.** *Let p be a prime number and $\nu_p(m)$ the exponent of p in the prime factorization of m. Then*

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

Proof. We have

$$\begin{aligned} s_p(n) &= \sum_{k<n} (s_p(k+1) - s_p(k)) \\ &= \sum_{k<n} (1 - (p-1)\nu_p(k+1)) \\ &= n - (p-1) \sum_{1 \leq k \leq n} \nu_p(k) \\ &= n - (p-1)\nu_p(n!). \end{aligned}$$

□

Applying Legendre's theorem three times, we obtain the interesting representation

$$s_p(n+t) - s_p(n) = s_p(t) - (p-1)\max\left\{k : p^k \mid \binom{n+t}{n}\right\}. \tag{2.3}$$

In combination with the theorem this says that for all k we have

$$\operatorname{dens}\left\{n : p^k \mid \binom{n+t}{n}\right\} = \operatorname{dens}\left\{n : p^k \mid \binom{n+t^R}{n}\right\},$$

which is a statement concerning columns in Pascal's triangle modulo powers of p.

Moreover, applying Kummer's theorem, we obtain the result that

$$\operatorname{dens}\{n : \operatorname{car}(n,t) = k\} = \operatorname{dens}\{n : \operatorname{car}(n,t^R) = k\},$$

where $\operatorname{car}(n,t)$ denotes the number of carries in the addition of n and t in base p.

Our research was motivated by a question of Thomas W. Cusick [14]: *Assume that t is a nonnegative integer and define*

$$c_t = \lim_{N\to\infty} \frac{1}{N} \#\{n < N : s_2(n+t) \geq s_2(n)\}. \tag{2.4}$$

Do we have $c_t > 1/2$ for all $t \geq 0$? In other words he asks whether for less than half of all n the binary sum of digits of $n+t$ is less than the sum of digits of n. The fact that this limit exists can be seen, for example, by showing that the sequence $n \mapsto \binom{n+t}{t}$ is ultimately periodic modulo p^k for all k, which is sufficient by the identity (2.3). To this end, we refer the reader to Zabek [56], who found the minimal period. We will prove the existence of the limit in a different way later on, see Lemma 2.6.

The inoccent looking question by Cusick does not seem to be easy to prove and to the author's knowledge it is still open (as of 2014). Howewer, we will be concerned with Cusick's question in chapter 3, where some partial answers will be given, among them a result that answers the question in the positive for t in a set of asymptotic density 1.

Cusick came up with this question while he was working on a combinatorial problem proposed by Tu and Deng [54] that is strongly related to Boolean functions with optimal cryptographic properties. In [15] some cases of this conjecture have been proved, and there are several other recent papers dealing with this subject, see for example [26, 25].

Although we could not give a complete answer to Cusick's original question, it follows from Theorem 2.2 that

$$c_t = c_{t^R} \tag{2.5}$$

for all $t \geq 0$.

In section 2.2.1 we will reprove the known result that the correlations γ_t for the 2-multiplicative function $n \mapsto \mathrm{e}(\vartheta s_2(n))$ satisfy the recurrence $\gamma_{2t} = \gamma_t$ and $\gamma_{2t+1} = \alpha\gamma_t + \beta\gamma_{t+1}$ for $\alpha = \mathrm{e}(\vartheta)/2$ and $\beta = \mathrm{e}(-\vartheta)/2$.

Motivated by this recurrence relation, we want to study more generally complex valued sequences z satisfying

$$\begin{aligned} z_{2t} &= z_t \\ z_{2t+1} &= \alpha z_t + \beta z_{t+1} \end{aligned} \tag{2.6}$$

for all $t \geq 1$, where $\alpha, \beta \in \mathbb{C}$ are constants. As it turns out, sequences of this kind satisfy the same kind of reflection property as in Theorem 2.1, which is our second main theorem.

Theorem 2.5. *Let α and β be complex numbers and $(z_n)_{n\geq 0}$ be a sequence satisfying the recurrence*

$$z_{2t} = z_t \quad \text{and} \quad z_{2t+1} = \alpha z_t + \beta z_{t+1}$$

for all $t \geq 1$. Then $z_{t^R} = z_t$, where the digit reversal is with respect to base 2.

We will give two (related) proofs of this theorem. Finally we note that it would be nice to find a common generalization of Theorems 2.5 and 2.1. The case that $q = 2$, $\alpha = \mathrm{e}(\vartheta)/2$ and $\beta = \mathrm{e}(-\vartheta)/2$ is contained as a special case in both theorems, and it is the only case covered by both, so that it seems natural to look for a more general statement: *let $q \geq 2$ and $\alpha_0, \dots, \alpha_{q-1}$ and $\beta_0, \dots, \beta_{q-1}$ be complex numbers. Assume that z is a sequence such that*

$$z_{qt+i} = \alpha_i z_t + \beta_i z_{t+1}$$

for $t \geq 0$ and $0 \leq i < q$. Under which conditions on the values α_i and β_i do we have

$$z_t = z_{t^R}?$$

We leave this question open.

2.2 Proofs

2.2.1 Auxiliary Lemmas

In order to prove Theorems 2.1 and 2.5, we want to derive the recurrence relation governing the correlation. As an essential tool for this we prove Lemma 2.6 below, which will also provide us with an alternative proof of the existence of the correlation of the q-multiplicative function $n \mapsto \mathrm{e}(\vartheta s_q(n))$. There is nothing new about these results, in fact we follow Bésineau [7] in order to obtain them. The following Lemma is essentially Lemme 1 in [7].

Lemma 2.6. *Let $q \geq 2$ and $t \geq 0$ be integers. There exists a partition $\mathcal{N}_{q,t}$ of the set of nonnegative integers having the properties that*

- *Each $N \in \mathcal{N}_{q,t}$ if of the form $a + q^k\mathbb{N}$, where $a < q^k$ and $k \geq 0$.*
- *The function $s_q(n+t) - s_q(n)$ is constant on each $N \in \mathcal{N}_{q,t}$.*
- *For all integers k the set $A_q(k,t) = \{n \in \mathbb{N} : s_q(n+t) - s_q(n) = k\}$ is a finite (possibly empty) union of elements of sets from the partition $\mathcal{N}_{q,t}$.*

In particular, each of the sets $A_q(k,t)$ possesses an asymptotic density $\delta_q(k,t)$. Moreover, the densities satisfy the following recurrence relation, for all k, $t \geq 0$ and $0 \leq b < q$:

$$\begin{aligned} \delta_q(k,0) &= \delta_{k,0} \text{ and} \\ \delta_q(k, qt+b) &= \frac{q-b}{q}\delta_q(k-b,t) + \frac{b}{q}\delta_q(k-b+q, t+1). \end{aligned} \tag{2.7}$$

Proof. For brevity, we set $d_q(n,t) = s_q(n+t) - s_q(n)$. For all n, $t \geq 0$ and $0 \leq a$, $b < q$, the values of d_q satisfy the property

$$d_q(qn+a, qt+b) = \begin{cases} d_q(n,t) + b & \text{if } a+b < q \\ d_q(n,t+1) + b - q & \text{if } a+b \geq q, \end{cases} \tag{2.8}$$

which follows easily from the elementary properties of the sum of digits: $s_q(qm) = s_q(m)$ and $s_q(qm+r) = s_q(m) + r$ for $0 \leq r < q$. We have $d_q(n,0) = 0$ for all n, therefore $A_q(k,0) = \mathbb{N}$ if $k = 0$ and $A_q(k,0) = \emptyset$ otherwise, which implies the first line of (2.7).

Moreover, using the representation of a natural number in base q, we obtain

$$d_q(n,1) = 1 - (q-1)\max\{k \geq 0 : q^k \mid n+1\}. \tag{2.9}$$

We prove the second line by induction on t.

In the case that $t = 1$ we have by (2.9) the decomposition

$$A_q(1-(q-1)r, t) = \bigcup_{s=1}^{q-1} \left(-1 + sq^r + q^{r+1}\mathbb{N}\right)$$

and $A_q(k,t) = \emptyset$ if k is not of the form $1-(q-1)r$. Let $t > 1$, $t = qm + b$. Equation (2.8) yields

$$A_q(k,t) = \bigcup_{s=0}^{q-b-1} (qA_q(k-b,m)+s) \cup \bigcup_{s=q-b}^{q-1} (qA_q(k-b+q,m+1)+s),$$

which is a disjoint union. This implies the second line of (2.7) and therefore the lemma is proved. □

We note (see [10, p. 22]) that this lemma can be generalized to arbitrary q-additive functions, which yields an alternative proof of Lemma 1.24. However, we do not follow this path.

Lemma 2.7. *Let $(a_k)_{k\in\mathbb{N}}$ be a bounded sequence of complex numbers. Assume that for all $z \in \mathbb{C}$ the limit*

$$\rho(z) := \lim_{N\to\infty} \frac{1}{N} |\{k < N : a_k = z\}|$$

exists. Then the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k<N} a_k$$

exists and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k<N} a_k = \sum_{z\in\mathbb{C}} \rho(z) z.$$

Proof. Without loss of generality let $|a_k| \leq 1$ for all $k \in \mathbb{N}$. We write $L = \{a_n : n \in \mathbb{N}\}$ and $\rho_N(z) = \frac{1}{N}|\{n < N : a_n = z\}|$ for abbreviation. Note that $\sum_{z\in L} \rho(z) = 1$ and $\sum_{z\in L} \rho_N(z) = 1$ for all N. Let $\varepsilon > 0$ be given and choose a finite set $A \subseteq L$ in such a way that $\left|\sum_{z\in L\setminus A} \rho(z)\right| < \varepsilon$ and M so large that $\frac{1}{N}|\{\rho_N(z) - \rho(z)\}| < \frac{\varepsilon}{|A|}$ for all $N \geq M$ and all $z \in A$. Then we have $\sum_{z\in L\setminus A} \rho_N(z) = 1 - \sum_{z\in A} \rho_N(z) \leq 1 - \sum_{z\in A} \rho(z) + |A|\frac{\varepsilon}{|A|} < 2\varepsilon$ and therefore $\left|\frac{1}{N}\sum_{n<N} a_n - \sum_{z\in L} \rho(z) z\right| = \left|\sum_{z\in L} (\rho_N(z) - \rho(z))\, z\right| \leq \left|\sum_{z\in A} (\rho_N(z) - \rho(z))\right| + \sum_{z\in L\setminus A} \rho_N(z) + \sum_{z\in L\setminus A} \rho(z) < 4\varepsilon$ for all $N \geq M$. □

Using Lemma 2.6 and Lemma 2.7 we get the following result (compare [7, Lemme 2]), which gives an alternative proof of the existence of the correlation.

Lemma 2.8. *Let ϑ be a real number and t a nonnegative integer. Let*

$$\delta(k,t) = \operatorname{dens}\{n : s_q(n+t) - s_q(n) = k\}.$$

Then

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n<N} e\left(\vartheta(s_q(n+t) - s_q(n))\right) = \sum_{k\in\mathbb{Z}} \delta(k,t) e\left(\vartheta k\right), \tag{2.10}$$

where the limit on the left hand side exists and the sum on the right hand side is absolutely convergent.

We define $\gamma_t(\vartheta)$ as the limit in (2.10). Formulas (2.7) and (2.10) imply the following result, which is formula (1) in [7].

Lemma 2.9. *Let $\vartheta \in \mathbb{R}$, $t \in \mathbb{N}$ and $0 \le b < q$. Then for all $\vartheta \in \mathbb{R}$, $t \in \mathbb{N}$ and $0 \le b < q$ the following recurrence holds:*

$$\begin{aligned} \gamma_0(\vartheta) &= 1 \\ \gamma_{qt+b}(\vartheta) &= \frac{q-b}{q} e\left(\vartheta b\right) \gamma_t(\vartheta) + \frac{b}{q} e\left(\vartheta(b-q)\right) \gamma_{t+1}(\vartheta) \text{ for } t \ge 0 \text{ and } 0 \le b < q. \end{aligned} \tag{2.11}$$

In particular, we have

$$\gamma_{qt}(\vartheta) = \gamma_t(\vartheta) \tag{2.12}$$

and

$$\gamma_1(\vartheta) = \frac{q-1}{qe(-\vartheta) - e(-\vartheta q)}. \tag{2.13}$$

For the Thue-Morse sequence ($q = 2$, $\vartheta = 1/2$) we get $\gamma_0 = 1$, $\gamma_1 = -1/3$, $\gamma_{2t} = \gamma_t$ and $\gamma_{2t+1} = -1/2(\gamma_t + \gamma_{t+1})$. For illustration, we list the first few values of the correlation of the Thue-Morse sequence.

t	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
γ_t	1	$-\frac{1}{3}$	$-\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{3}$	0	$\frac{1}{3}$	0	$-\frac{1}{3}$	$\frac{1}{6}$	0	$-\frac{1}{6}$	$\frac{1}{3}$	$-\frac{1}{6}$	0	$\frac{1}{6}$

t	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
γ_t	$-\frac{1}{3}$	$\frac{1}{12}$	$\frac{1}{6}$	$-\frac{1}{12}$	0	$\frac{1}{12}$	$-\frac{1}{6}$	$-\frac{1}{12}$	$\frac{1}{3}$	$-\frac{1}{12}$	$-\frac{1}{6}$	$\frac{1}{12}$	0	$-\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$

As we noted before, the sum-of-digits case in base $q = 2$ is covered by both Theorem 2.1 and Theorem 2.5. For $q > 2$ equation (2.11) gives a different recurrence relation. Nevertheless, the correlations $\gamma_t(\vartheta)$ of the sum-of-digits function satisfy the same kind of reflection property as in the case $q = 2$.

2.2.2 Proof of the reflection property for the sum of digits

Proof of Theorem 2.1. By Lemma 2.9 the correlations $\gamma_t(\vartheta)$ satisfy the properties that $\gamma_0(\vartheta) = 1$ and

$$\gamma_{qt+k}(\vartheta) = \frac{q-k}{q} \operatorname{e}(\vartheta k)\gamma_t(\vartheta) + \frac{k}{q} \operatorname{e}(-\vartheta(q-k))\gamma_{t+1}(\vartheta)$$

for $t \geq 0$ and $0 \leq k < q$. In particular, we have $\gamma_{qt}(\vartheta) = \gamma_t(\vartheta)$ and $u := \gamma_1(\vartheta) = (q-1)/(q\,\mathrm{e}(-\vartheta) - \mathrm{e}(-\vartheta q))$. It is not difficult to see that we can represent $\gamma_t(\vartheta)$ with the help of transition matrices. Set

$$A(k) = \begin{pmatrix} \frac{q-k}{q}\,\mathrm{e}(\vartheta k) & \frac{k}{q}\,\mathrm{e}(-\vartheta(q-k)) \\ \frac{q-k-1}{q}\,\mathrm{e}(\vartheta(k+1)) & \frac{k+1}{q}\,\mathrm{e}(-\vartheta(q-k-1)) \end{pmatrix}.$$

Then we have

$$\gamma_t(\vartheta) = (1\,,0)\,A(\varepsilon_0(t))\cdots A(\varepsilon_\nu(t))\begin{pmatrix}1\\u\end{pmatrix}. \tag{2.14}$$

Note that it is not important whether the proper representation of t is used in order to calculate $\gamma_t(\vartheta)$. That is, we do not have to claim $\varepsilon_\nu(t) \neq 0$. This follows from the fact that $(1\,,u)^T$ is a right eigenvector of $A(0)$ to the eigenvalue 1. Note furthermore that $\gamma_{qt}(\vartheta) = \gamma_t(\vartheta)$ corresponds to the fact that $(1\,,0)$ is a left eigenvector of $A(0)$ to the eigenvalue 1. Set

$$S = \begin{pmatrix} 1 & \bar{u} \\ 0 & 1 \end{pmatrix}.$$

Proposition 2.10. *Let $\ell \geq 0$ and $(\varepsilon_0, \dots, \varepsilon_\ell) \in \{0, \dots, q-1\}^{\ell+1}$. Then we have*

$$(1\,,0)\,S^{-1}A(\varepsilon_0)\cdots A(\varepsilon_\ell)\begin{pmatrix}1\\u\end{pmatrix} = (1\,,0)\,A(\varepsilon_\ell)\cdots A(\varepsilon_0)S\begin{pmatrix}1-|u|^2\\0\end{pmatrix} \tag{2.15}$$

and

$$(0\,,\bar{u})\,S^{-1}A(\varepsilon_0)\cdots A(\varepsilon_\ell)\begin{pmatrix}1\\u\end{pmatrix} = (1\,,0)\,A(\varepsilon_\ell)\cdots A(\varepsilon_0)S\begin{pmatrix}0\\u\end{pmatrix}. \tag{2.16}$$

This proposition immediately implies Theorem 2.1. Indeed, if we sum up (2.15) and (2.16) we obtain

$$(1\,,\bar{u})\,S^{-1}A(\varepsilon_0)\cdots A(\varepsilon_\ell)\begin{pmatrix}1\\u\end{pmatrix} = (1\,,0)\,A(\varepsilon_\ell)\cdots A(\varepsilon_0)S\begin{pmatrix}1-|u|^2\\u\end{pmatrix}.$$

Since $(1\,,\bar{u})S^{-1} = (1\,,0)$ and $S(1-|u|^2\,,u)^T = (1\,,u)^T$, relation (2.14) implies that $\gamma_t(\vartheta) = \gamma_{t^R}(\vartheta)$. □

Proof of Proposition 2.10. We will show this result by induction on ℓ. For notational convenience we set

$$A(\varepsilon) = \begin{pmatrix} a_1(\varepsilon) & a_2(\varepsilon) \\ a_3(\varepsilon) & a_4(\varepsilon) \end{pmatrix} \qquad \text{and} \qquad S^{-1}A(\varepsilon)S = \begin{pmatrix} s_1(\varepsilon) & s_2(\varepsilon) \\ s_3(\varepsilon) & s_4(\varepsilon) \end{pmatrix}.$$

Throughout the proof, we will use (at several places) the relation

$$a_1(\varepsilon)|u|^2 + a_2(\varepsilon)u = a_3(\varepsilon)\bar{u} + a_4(\varepsilon)|u|^2 \tag{2.17}$$

which holds for $0 \le \varepsilon < q$. The validity of (2.17) is easily seen by multiplying both sides by $|u|^{-2}$ and evaluating them: The left hand side gives

$$\frac{e(\vartheta\varepsilon)}{q(q-1)}\left((q-1)(q-\varepsilon)+\varepsilon e(-\vartheta q)(qe(\vartheta)-e(\vartheta q))\right)$$

which simplifies to

$$\frac{e(\vartheta\varepsilon)}{q-1}\left(q-\varepsilon-1+\varepsilon e(-\vartheta(q-1))\right),$$

while the right hand side gives

$$\frac{e(\vartheta\varepsilon)}{q(q-1)}\left((q-\varepsilon-1)e(\vartheta)(qe(-\vartheta)-e(-\vartheta q))+(q-1)(\varepsilon+1)e(-\vartheta(q-1))\right)$$

which evaluates to the same expression.

If $\ell = 0$ we have to show that

$$(1\,,0)\,S^{-1}A(\varepsilon_0)\begin{pmatrix}1\\ u\end{pmatrix} = (1\,,0)\,A(\varepsilon_0)S\begin{pmatrix}1-|u|^2\\ 0\end{pmatrix} \tag{2.18}$$

and

$$(0\,,\bar{u})\,S^{-1}A(\varepsilon_0)\begin{pmatrix}1\\ u\end{pmatrix} = (1\,,0)\,A(\varepsilon_0)S\begin{pmatrix}0\\ u\end{pmatrix}. \tag{2.19}$$

Equation (2.18) is satisfied if $a_1(\varepsilon_0) + a_2(\varepsilon_0)u - a_3(\varepsilon_0)\bar{u} - a_4(\varepsilon_0)|u|^2 = a_1(\varepsilon_0)(1-|u|^2)$. Using (2.17), we see that this holds true indeed. Equation (2.19) is also equivalent to (2.17) and we are done. Assume now that $\ell \ge 1$. Set

$$\begin{pmatrix}\mathfrak{a}\\ \mathfrak{b}\end{pmatrix} = S^{-1}A(\varepsilon_1)\dots A(\varepsilon_\ell)\begin{pmatrix}1\\ u\end{pmatrix} \qquad \text{and} \qquad (\mathfrak{a}'\,,\mathfrak{b}') = (1\,,0)A(\varepsilon_\ell)\cdots A(\varepsilon_1)S.$$

The induction hypothesis implies that

$$\mathfrak{a} = \mathfrak{a}'(1-|u|^2) \qquad \text{and} \qquad \mathfrak{b}\bar{u} = \mathfrak{b}'u. \tag{2.20}$$

In order to prove (2.15), we have to show that

$$(1,\,0)S^{-1}A(\varepsilon_0)S\begin{pmatrix}\mathfrak{a}\\ \mathfrak{b}\end{pmatrix} = (\mathfrak{a}'\,,\mathfrak{b}')S^{-1}A(\varepsilon_0)S\begin{pmatrix}1-|u|^2\\ 0\end{pmatrix}. \tag{2.21}$$

This is equivalent to $s_1(\varepsilon_0)\mathfrak{a} + s_2(\varepsilon_0)\mathfrak{b} = s_1(\varepsilon_0)(1-|u|^2)\mathfrak{a}' + s_3(\varepsilon_0)(1-|u|^2)\mathfrak{b}'$. Using (2.20), we see that this holds true if $s_2(\varepsilon_0)u/\bar{u} = s_3(\varepsilon_0)(1-|u|^2)$. Note that $s_2(\varepsilon_0)$ and $s_3(\varepsilon_0)$ are given by $s_2(\varepsilon_0) = a_1(\varepsilon_0)\bar{u} + a_2(\varepsilon_0) - \bar{u}^2a_3(\varepsilon_0) - \bar{u}a_4(\varepsilon_0)$ and $s_3(\varepsilon_0) = a_3(\varepsilon_0)$. Using these relations and (2.17), we see that (2.21) holds true. The validity of (2.16) can be shown the same way. This finally proves Proposition 2.10. □

To the same extent to which the reflection property for the sum of digits is unexpected, the proof is unintuitive. So far it has failed to lead us to deeper insight into the problem and the author still does not "understand" the real reason for the reflection property to hold.

Proof of Theorem 2.2. Using the dominated convergence theorem, we see that

$$\begin{aligned}\delta(k,t) &= \lim_{x\to\infty}\frac{1}{x}\#\{n<x : s_q(n+t)-s_q(n)=k\}\\ &= \lim_{x\to\infty}\frac{1}{x}\sum_{n<x}\int_0^1 \mathrm{e}(\vartheta(s_q(n+t)-s_q(n)-k))\,\mathrm{d}\vartheta\\ &= \int_0^1 \lim_{x\to\infty}\sum_{n<x}\frac{1}{x}\,\mathrm{e}(\vartheta(s_q(n+t)-s_q(n)-k))\,\mathrm{d}\vartheta.\end{aligned}$$

Thus we have

$$\delta(k,t) = \int_0^1 \gamma_t(\vartheta)\,\mathrm{e}(-\vartheta k)\,\mathrm{d}\vartheta. \tag{2.22}$$

By Theorem 2.1 we have $\gamma_t(\vartheta)=\gamma_{t^R}(\vartheta)$ and we get $\delta(k,t)=\delta(k,t^R)$. □

It remains to prove (2.5). Equation (2.3) implies $s_q(n+t)-s_q(n)\le s_q(t)$. Therefore we have $c_t=\sum_{k=0}^{s_q(t)}\delta(k,t)$. Since $\delta(k,t)=\delta\left(k,t^R\right)$, we are done.

2.2.3 A remark on the integral representation of $\delta(k,t)$

From the integral representation (2.22) we also get the following formula for c_t.

Corollary 2.11. *Let $t\ge 0$. We have*

$$c_t = \frac{1}{2}+\int_0^1 \mathrm{Re}\,\frac{\gamma_t(\vartheta)}{1-\mathrm{e}(-\vartheta)}\,\mathrm{d}\vartheta.$$

The statement $c_t>1/2$ is therefore equivalent to showing positivity of an integral. An approach to prove such a statement is to find a property of the function $\vartheta\mapsto\gamma_t(\vartheta)$ that is preserved under the recurrence relation governing the correlation and from which we can prove positivity of the integral. So far we have not succeeded to find such a property, however.

The proof of Corollary 2.11 uses the geometric sum formula and the following identity.

Lemma 2.12. *For $k\ge 1$ we have*

$$\int_0^1 \sin(2\pi k\vartheta)\cot(\pi\vartheta)\,\mathrm{d}\vartheta = 1, \tag{2.23}$$

where the integrand is bounded on $(0,1)$.

Proof. We have

$$\begin{aligned}&\sin(2\pi k\vartheta)\cot(\pi\vartheta)\\ =\;& \sin(2\pi(k-1)\vartheta)\cos(2\pi\vartheta)\cot(\pi\vartheta)+\cos(2\pi(k-1)\vartheta)\sin(2\pi\vartheta)\cot(\pi\vartheta)\\ =\;& \sin(2\pi(k-1)\vartheta)\left(1-2\sin^2(\pi\vartheta)\right)\frac{\cos(\pi\vartheta)}{\sin(\pi\vartheta)}+2\cos(2\pi(k-1)\vartheta)\sin(\pi\vartheta)\cos(\pi\vartheta)\frac{\cos(\pi\vartheta)}{\sin(\pi\vartheta)}\\ =\;& \sin(2\pi(k-1)\vartheta)\cot(\pi\vartheta)-\sin(2\pi(k-1)\vartheta)\sin(2\pi\vartheta)+\cos(2\pi(k-1)\vartheta)\left(\cos(\pi\vartheta)+1\right)\\ =\;& \sin(2\pi(k-1)\vartheta)\cot(\pi\vartheta)+\cos(2\pi(k-1/2)\vartheta)+\cos(2\pi(k-1)\vartheta).\end{aligned}$$

Assume first that $k = 1$. The first summand is identically zero on $(0, 1)$, the integral from 0 to 1 of the second summand equals zero, and the third summand is identically 1. This implies the statement for $k = 1$. For $k \geq 2$ the first summand is bounded by the induction hypothesis and its contribution to the integral is 1. The other summands contribute nothing to the integral. The statement is therefore proved. □

Proof of Corollary 2.11. We note that

$$\gamma_t(\vartheta) = \sum_{k<m} \delta(k,t)\,\mathrm{e}(k\vartheta)$$

for some $m \geq 1$, which can be shown by induction easily. Necessarily we have $\delta(k,t) = 0$ for $k \geq m$. It follows from (2.22) that

$$\begin{aligned}
c_t &= \sum_{0\leq k<m} \delta(k,t) \\
&= \int_0^1 \gamma_t(\vartheta) \sum_{0\leq k<m} \mathrm{e}(-k\vartheta)\,\mathrm{d}\vartheta \\
&= \int_0^1 \gamma_t(\vartheta) \frac{1-\mathrm{e}(-m\vartheta)}{1-\mathrm{e}(-\vartheta)}\,\mathrm{d}\vartheta \\
&= \int_0^1 \operatorname{Re}\frac{\gamma_t(\vartheta)}{1-\mathrm{e}(-\vartheta)} - \operatorname{Re}\frac{\gamma_t(\vartheta)\,\mathrm{e}(-m\vartheta)}{1-\mathrm{e}(-\vartheta)}\,\mathrm{d}\vartheta \\
&= \int_0^1 \operatorname{Re}\frac{\gamma_t(\vartheta)}{1-\mathrm{e}(-\vartheta)} - \frac{1}{2}\operatorname{Re}\left(\gamma_t(\vartheta)\,\mathrm{e}(-m\vartheta)\right) + \frac{1}{2}\operatorname{Im}\left(\gamma_t(\vartheta)\,\mathrm{e}(-m\vartheta)\right)\cot(-\pi\vartheta)\,\mathrm{d}\vartheta,
\end{aligned}$$

where we have used the formulas

$$\operatorname{Re}\frac{1}{1-\mathrm{e}(-\vartheta)} = \frac{1}{2}$$

and

$$\operatorname{Im}\frac{1}{1-\mathrm{e}(-\vartheta)} = \frac{1}{2}\cot(-\pi\vartheta).$$

Since $\sum_{k<m}\delta(k,t) = 1$, it follows that

$$\gamma_t(\vartheta)\,\mathrm{e}(-m\vartheta) = \sum_{\ell\geq 1} a_\ell\,\mathrm{e}(-\ell\vartheta)$$

for some nonnegative a_ℓ such that $\sum_{\ell\geq 1} a_\ell = 1$.

Since m is large enough, the integral over the second summand is zero. We obtain

$$c_t = \int_0^1 \operatorname{Re}\frac{\gamma_t(\vartheta)}{1-\mathrm{e}(-\vartheta)} + \frac{1}{2}\sum_{\ell\geq 1} a_\ell \sin(-2\pi\ell\vartheta)\cot(-\pi\vartheta)\,\mathrm{d}\vartheta.$$

The second summand is a bounded function, therefore the statement follows by an application of the identity (2.23). □

2.2.4 First proof of Theorem 2.5

Clearly $z\left(0^R\right) = z\left(0\right)$. If $z\left(1\right) = 0$, then $z(k) = 0$ for $k \geq 1$, therefore the assertion is trivial in this case. Multiplication of the sequence with a constant preserves the recurrence, therefore we may assume without loss of generality that $z\left(1\right) = 1$.

The main argument, which represents the induction step in the proof of the theorem, is the following lemma.

Lemma 2.13. *Let*

$$A = \begin{pmatrix} 1 & 0 \\ \alpha & \beta \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} \alpha & \beta \\ 0 & 1 \end{pmatrix}.$$

Then

$$\begin{aligned}
\begin{pmatrix} \alpha & \beta \end{pmatrix} AA &= (-\beta) \begin{pmatrix} \alpha & \beta \end{pmatrix} + (\beta+1) \begin{pmatrix} \alpha & \beta \end{pmatrix} A, \\
\begin{pmatrix} 1 & 1 \end{pmatrix} {}^tA{}^tA &= (-\beta) \begin{pmatrix} 1 & 1 \end{pmatrix} + (\beta+1) \begin{pmatrix} 1 & 1 \end{pmatrix} {}^tA, \\
\\
\begin{pmatrix} \alpha & \beta \end{pmatrix} AB &= \alpha \begin{pmatrix} \alpha & \beta \end{pmatrix} + \beta \begin{pmatrix} \alpha & \beta \end{pmatrix} B, \\
\begin{pmatrix} 1 & 1 \end{pmatrix} {}^tA{}^tB &= \alpha \begin{pmatrix} 1 & 1 \end{pmatrix} + \beta \begin{pmatrix} 1 & 1 \end{pmatrix} {}^tB, \\
\\
\begin{pmatrix} \alpha & \beta \end{pmatrix} BA &= \beta \begin{pmatrix} \alpha & \beta \end{pmatrix} + \alpha \begin{pmatrix} \alpha & \beta \end{pmatrix} A, \\
\begin{pmatrix} 1 & 1 \end{pmatrix} {}^tB{}^tA &= \beta \begin{pmatrix} 1 & 1 \end{pmatrix} + \alpha \begin{pmatrix} 1 & 1 \end{pmatrix} {}^tA, \\
\\
\begin{pmatrix} \alpha & \beta \end{pmatrix} BB &= (-\alpha) \begin{pmatrix} \alpha & \beta \end{pmatrix} + (\alpha+1) \begin{pmatrix} \alpha & \beta \end{pmatrix} B, \\
\begin{pmatrix} 1 & 1 \end{pmatrix} {}^tB{}^tB &= (-\alpha) \begin{pmatrix} 1 & 1 \end{pmatrix} + (\alpha+1) \begin{pmatrix} 1 & 1 \end{pmatrix} {}^tB.
\end{aligned}$$

The proof is a simple, but slightly tedious calculation. We skip it.

Let $t \geq 1$ be an odd integer. The general case follows from this one by repeatedly using the relation $z_{2t} = z_t$. Let $t = \sum_{i \leq \nu} \varepsilon_i 2^i$ be the binary representation of t and $\varepsilon_\nu \neq 0$. We prove the theorem by induction on ν. The case that $\nu \leq 1$ is trivial, since in this case we have $t^R = t$.

We write $A(0) = \left(\begin{smallmatrix} 1 & 0 \\ \alpha & \beta \end{smallmatrix}\right)$ and $A(1) = \left(\begin{smallmatrix} \alpha & \beta \\ 0 & 1 \end{smallmatrix}\right)$. By a simple application of the relations $z_{2t} = z_t$ and $z_{2t+1} = \alpha z_t + \beta z_{t+1}$ we have $\left(\begin{smallmatrix} z(2s) \\ z(2s+1) \end{smallmatrix}\right) = A(0) \left(\begin{smallmatrix} z(s) \\ z(s+1) \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} z(2s+1) \\ z(2s+2) \end{smallmatrix}\right) = A(1) \left(\begin{smallmatrix} z(s) \\ z(s+1) \end{smallmatrix}\right)$.

Since t is odd and $z(1) = z(2) = 1$ from these identities it follows by induction that

$$z(t) = \begin{pmatrix} \alpha & \beta \end{pmatrix} A(\varepsilon_1) \cdots A(\varepsilon_{\nu-1}) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{2.24}$$

and the statement of the theorem is equivalent to the assertion that

$$\begin{pmatrix} \alpha & \beta \end{pmatrix} A(\varepsilon_1) \cdots A(\varepsilon_{\nu-1}) \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \end{pmatrix} {}^tA(\varepsilon_1) \cdots {}^tA(\varepsilon_{\nu-1}) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \tag{2.25}$$

for all $\nu \geq 1$ and all finite sequences $(\varepsilon_1, \ldots, \varepsilon_{\nu-1})$ in $\{0, 1\}$. We prove the statement by induction on ν, using Lemma 2.13. The statement is obvious for $\nu \leq 2$.

For $\nu > 2$ we have four cases, corresponding to the four possible values of $(\varepsilon_1, \varepsilon_2)$. By Lemma 2.13 there exist in each of the four cases coefficients x and y such that

$$\begin{pmatrix} \alpha & \beta \end{pmatrix} A(\varepsilon_1) A(\varepsilon_2) = x \begin{pmatrix} \alpha & \beta \end{pmatrix} + y \begin{pmatrix} \alpha & \beta \end{pmatrix} A(\varepsilon_2)$$

and

$$\begin{pmatrix}1 & 1\end{pmatrix} {}^tA(\varepsilon_1){}^tA(\varepsilon_2) = x \begin{pmatrix}1 & 1\end{pmatrix} + y \begin{pmatrix}1 & 1\end{pmatrix} {}^tA(\varepsilon_2).$$

Applying the induction hypothesis (2.25) to the sequences $(\varepsilon_2, \ldots, \varepsilon_{\nu-1})$ and $(\varepsilon_3, \ldots, \varepsilon_{\nu-1})$ we obtain the statement.

2.2.5 Second proof of Theorem 2.5

The idea of proof of the second proof of the reflection property is taken from Urbiha [55, Lemma 3]. In this article he studies old and new properties of the function D defined by $D(0) = 0$, $D(1) = 1$, $D(2n) = D(n)$ and $D(2n+1) = D(n) + D(n+1)$, which corresponds to the case that $\alpha = \beta = 1$. Among the properties he (re)proved is the reflection property for this special case.

Lemma 2.14. *Let $k \geq 1$ be an integer and $\alpha, \beta \in \mathbb{R}$. Then*

$$\begin{pmatrix}\alpha & \beta\\ 0 & 1\end{pmatrix} \begin{pmatrix}1 & 0\\ \alpha & \beta\end{pmatrix}^k \begin{pmatrix}1\\ 1\end{pmatrix} = \beta \begin{pmatrix}\alpha & \beta\\ 0 & 1\end{pmatrix} \begin{pmatrix}1 & 0\\ \alpha & \beta\end{pmatrix}^{k-1} \begin{pmatrix}1\\ 1\end{pmatrix} + \alpha \begin{pmatrix}1\\ 1\end{pmatrix}$$

and

$$\begin{pmatrix}1 & 0\\ \alpha & \beta\end{pmatrix} \begin{pmatrix}\alpha & \beta\\ 0 & 1\end{pmatrix}^k \begin{pmatrix}1\\ 1\end{pmatrix} = \alpha \begin{pmatrix}1 & 0\\ \alpha & \beta\end{pmatrix} \begin{pmatrix}\alpha & \beta\\ 0 & 1\end{pmatrix}^{k-1} \begin{pmatrix}1\\ 1\end{pmatrix} + \beta \begin{pmatrix}1\\ 1\end{pmatrix}$$

Proof. The proof of this statement is simple, noting that

$$\begin{pmatrix}\alpha & \beta\\ 0 & 1\end{pmatrix}^s = \begin{pmatrix}\alpha^s & \beta \sum_{i<s} \alpha^i\\ 0 & 1\end{pmatrix}$$

and

$$\begin{pmatrix}1 & 0\\ \alpha & \beta\end{pmatrix}^s = \begin{pmatrix}1 & 0\\ \alpha \sum_{i<s} \beta^i & \beta^s\end{pmatrix}$$

for $s \geq 0$. □

Let $a_i \in \{0, 1\}$ for $1 \leq i < \nu$. We prove by induction on $\nu \geq 1$ that

$$z\left((1a_1 \cdots a_{\nu-1}1)_2\right) = z\left((1a_{\nu-1} \cdots a_1 1)_2\right).$$

The statement is trivial for $\nu \leq 2$. Let $\nu \geq 3$ and assume that $a_1 = 0$. If $a_2 = \cdots = a_{\nu-1} = 0$, there is nothing to prove. Otherwise there is some $k \geq 1$ such that $a_1 = \cdots = a_k = 0$ and $a_{k+1} = 1$.

By (2.24), the first point of Lemma 2.14, the induction hypothesis and the relations

$z_{2t} = z_t$ and $z_{2t+1} = \alpha z_t + \beta z_{t+1}$ we have

$$\begin{aligned} z\left((10a_2\cdots a_{\nu-1}1)_2\right) &= \begin{pmatrix}\alpha & \beta\end{pmatrix} A(a_{\nu-1})\cdots A(a_{k+2})A(1)A(0)^k \begin{pmatrix}1\\1\end{pmatrix} \\ &= \beta \begin{pmatrix}\alpha & \beta\end{pmatrix} A(a_{\nu-1})\cdots A(a_{k+2})A(1)A(0)^{k-1} \begin{pmatrix}1\\1\end{pmatrix} \\ &+ \alpha \begin{pmatrix}\alpha & \beta\end{pmatrix} A(a_{\nu-1})\cdots A(a_{k+2}) \begin{pmatrix}1\\1\end{pmatrix} \\ &= \alpha z\left((a_{k+1}\cdots a_{\nu-1}1)_2\right) + \beta z\left((1a_2\cdots a_{\nu-1}1)_2\right) \\ &= \alpha z\left((1a_{\nu-1}\cdots a_20)_2\right) + \beta z\left((1a_{\nu-1}\cdots a_21)_2\right) \\ &= z\left((1a_{\nu-1}\cdots a_201)_2\right). \end{aligned}$$

Now assume that $a_1 = 1$. The case that $a_2 = \cdots = a_{\nu-1} = 1$ is trivial. Let $k \geq 1$ be such that $a_1 = \ldots = a_k = 1$ and $a_{k+1} = 0$. Similarly to the calculation above we have

$$\begin{aligned} z\left((11a_2\cdots a_{\nu-1}1)_2\right) &= \begin{pmatrix}\alpha & \beta\end{pmatrix} A(a_{\nu-1})\cdots A(a_{k+2})A(0)A(1)^k \begin{pmatrix}1\\1\end{pmatrix} \\ &= \alpha \begin{pmatrix}\alpha & \beta\end{pmatrix} A(a_{\nu-1})\cdots A(a_{k+2})A(0)A(1)^{k-1} \begin{pmatrix}1\\1\end{pmatrix} \\ &+ \beta \begin{pmatrix}\alpha & \beta\end{pmatrix} A(a_{\nu-1})\cdots A(a_{k+2}) \begin{pmatrix}1\\1\end{pmatrix} \\ &= \alpha z\left((a_1\cdots a_{\nu-1}1)_2\right) + \beta z\left((1a_{k+2}\cdots a_{\nu-1}1)_2\right) \\ &= \alpha z\left(\left(1a_{\nu-1}\cdots a_{k+2}01^k\right)_2\right) + \beta z\left(\left(1a_{\nu-1}\cdots a_{k+2}10^k\right)_2\right) \\ &= z\left((1a_{\nu-1}\cdots a_211)_2\right). \end{aligned}$$

This finishes the second proof of the curious reflection property.

2.3 Facts on correlations

In this section we are still concerned with sequences z such that $z_{2t} = z_t$ and $z_{2t+1} = \alpha z_t + \beta z_{t+1}$. What follows is a collection of small observations that we encountered in our quest of understanding such sequences.

First we generalize (in the case that $q = 2$) the recurrence relation for the correlations to the case of the more general expression

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n<N} \mathrm{e}\left(\vartheta_0 s_2(n) + \vartheta_1 s_2(n+1) + \cdots + \vartheta_K s_2(n+K)\right).$$

Second, we establish a link between sequences z as above and determinants of tridiagonal matrices, so-called continuants.

Next we study the square mean of the correlation of the Thue-Morse sequence, exhibiting an explicit formula for the quantity

$$\frac{1}{2^k} \sum_{2^k \leq t < 2^{k+1}} |\gamma_t|^2.$$

Note that this quantity converges to 0 by Corollary 1.27. The explicit formula gives the rate of convergence to zero, that is, a measure of pseudorandomness of the Thue-Morse sequence.

Fourth, we study the discrete Fourier transform of the correlation γ (on dyadic intervals $[2^\lambda, \dots, 2^{\lambda+1}-1]$) of the 2-multiplicative function $n \mapsto \mathrm{e}(\vartheta s_2(n))$. By a short calculation we obtain an explicit formula for the Fourier coefficients.

Afterwards we consider the ordinary generating function of the correlation function γ and derive an explicit representation.

Finally, we give a graphical representation of the recurrence relation $z_{2t} = z_t$, $z_{2t+1} = \alpha z_t + \beta z_{t+1}$.

Generalized correlations

As a generalization of the (auto)correlation

$$\gamma_t(\vartheta) = \lim_{N\to\infty} \frac{1}{N} \sum_{n<N} \mathrm{e}\left(\vartheta(s_2(n+t) - s_2(n))\right)$$

we study *generalized correlations*. For simplicity, we only treat the case of the binary sum-of-digits function.

Let $(a_k)_{k\ge 0}$ be a sequence of real numbers such that almost all $a_k = 0$. Set

$$S_N(\vartheta_0, \dots, \vartheta_K) = \sum_{n<N} \mathrm{e}\left(\vartheta_0 s_2(n+0) + \cdots + \vartheta_K s_2(n+K)\right).$$

Then by a straightforward calculation we get

$$\begin{aligned}
S_{2N}(\vartheta_0, \dots, \vartheta_{2k}) &= \sum_{n<2N} \mathrm{e}\left(\vartheta_0 s_2(n+0) + \cdots + \vartheta_{2k} s_2(n+2k)\right) \\
&= \mathrm{e}\left(\vartheta_1 + \vartheta_3 + \ldots + \vartheta_{2k-1}\right) \sum_{n<N} \mathrm{e}\left((\vartheta_0+\vartheta_1)s_2(n+0) + \cdots \right. \\
&\qquad \left. + (\vartheta_{2k-2} + \vartheta_{2k-1}) s_2(n+k-1) + \vartheta_{2k} s_2(n+k)\right) \\
&\quad + \mathrm{e}\left(\vartheta_0 + \vartheta_2 + \ldots + \vartheta_{2k}\right) \sum_{n<N} \mathrm{e}\left(\vartheta_0 s_2(n+0) + (\vartheta_1+\vartheta_2)s_2(n+1) + \cdots \right. \\
&\qquad \left. + (\vartheta_{2k-1} + \vartheta_{2k}) s_2(n+k)\right) \\
&= \mathrm{e}\left(\vartheta_1 + \vartheta_3 + \cdots + \vartheta_{2k-1}\right) S_N(\vartheta_0 + \vartheta_1, \dots, \vartheta_{2k-2} + \vartheta_{2k-1}, \vartheta_{2k}) \\
&\qquad + \mathrm{e}\left(\vartheta_0 + \vartheta_2 + \cdots + \vartheta_{2k}\right) S_N(\vartheta_0, \vartheta_1 + \vartheta_2, \dots, \vartheta_{2k-1} + \vartheta_{2k})
\end{aligned}$$

By induction on L it follows that for all families $(\vartheta_0, \dots, \vartheta_{2^L})$ the limit

$$\gamma(\vartheta_0, \dots, \vartheta_{2^L}) = \lim_{N\to\infty} \frac{1}{N} S_N(\vartheta_0, \dots, \vartheta_{2^L})$$

exists, the case $L = 1$ being just the existence of the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n<N} \mathrm{e}\left(\vartheta_0 s(n) + \vartheta_1 s(n+1)\right).$$

By noting that $S_N(\vartheta_0,\dots,\vartheta_k) = S_N(\vartheta_0,\dots,\vartheta_k,0)$ we may extend the above definition by setting

$$\gamma(\vartheta_0,\dots,\vartheta_k) = \lim_{N\to\infty}\frac{1}{N}S_N(\vartheta_0,\dots,\vartheta_k)$$

for an arbitrary integer $k \geq 0$. These correlations satisfy the recurrence relation from above, viz.

$$\begin{aligned}&\gamma(\vartheta_0,\dots,\vartheta_{2k})\\ &\quad= \tfrac{1}{2}\,\mathrm{e}(\vartheta_1+\vartheta_3+\cdots+\vartheta_{2k-1})\,\gamma(\vartheta_0+\vartheta_1,\quad\dots\quad,\quad\dots\quad,\vartheta_{2k-2}+\vartheta_{2k-1},\vartheta_{2k}\qquad)\\ &\quad+ \tfrac{1}{2}\,\mathrm{e}(\vartheta_0+\vartheta_2+\cdots+\vartheta_{2k}\quad)\,\gamma(\vartheta_0\qquad,\vartheta_1+\vartheta_2,\quad\dots\quad,\qquad\dots\qquad,\vartheta_{2k-1}+\vartheta_{2k}).\end{aligned}$$

The correlations can be used to compute the frequency with which a certain subblock of the Thue-Morse sequence appears. Let $=(a_0,\dots,a_k) \in \{0,1\}^{k+1}$. We have

$$\begin{aligned}&|\{n<N : (t_{n+0},\dots,t_{n+k}) = (a_0,\dots,a_k)\}|\\ &\qquad= \frac{1}{2^{k+1}}\sum_{\mathbf{h}\in\{0,1\}^{k+1}}\sum_{n<N}\mathrm{e}\left(h_0\frac{a_0-s_2(n+0)}{2}+\cdots+h_k\frac{a_k-s_2(n+k)}{2}\right)\\ &\qquad= \frac{1}{2^{k+1}}\sum_{\mathbf{h}\in\{0,1\}^{k+1}}(-1)^{\mathbf{h}\cdot\mathbf{a}}S_N\left(h_0/2,\dots,h_k/2\right), \quad (2.26)\end{aligned}$$

from which a corresponding formula for the subblock density follows:

$$\operatorname{dens}\{n : (t_{n+0},\dots,t_{n+k}) = (a_0,\dots,a_k)\} = \frac{1}{2^{k+1}}\sum_{\mathbf{h}\in\{0,1\}^{k+1}}(-1)^{\mathbf{h}\cdot\mathbf{a}}\gamma\left(h_0/2,\dots,h_k/2\right).$$

As an example, we compute the densities of the substrings of length 5 of the Thue-Morse sequence.

We note that $\gamma(\frac{1}{2}\mathbf{a}) = 0$ if $\mathbf{a}$ has an odd number of ones, which can be seen by an easy induction.

For brevity, we write $t_0t_1\cdots t_{k-1}$ instead of $(t_0,t_1,\dots,t_{k-1})$.

From the recurrence relation and the initial correlation we get the following tables.

$\mathbf{t}$	00	11
$\gamma(\frac{1}{2}\mathbf{t})$	1	$-\frac{1}{3}$

$\mathbf{t}$	000	011	101	110
$\gamma(\frac{1}{2}\mathbf{t})$	1	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$

$\mathbf{t}$	00000	00011	00101	00110	01001	01010	01100	01111
$\gamma(\frac{1}{2}\mathbf{t})$	1	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	$\frac{1}{3}$

$\mathbf{t}$	10001	10010	10100	10111	11000	11011	11101	11110
$\gamma(\frac{1}{2}\mathbf{t})$	$-\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Feeding these values into the formula for the densities, we obtain that each of the blocks 00101, 00110, 01001, 01011, 01100, 01101, 10010, 10011, 10100, 10110, 11001 and 11010 occurs with the frequency $\frac{1}{12}$.

Continuants

It might be interesting to note that sequences z such that $z_{2t} = z_t$ and $z_{2t+1} = \alpha z_t + \beta z_{t+1}$ can be related to determinants. In the theory of continued fractions there is the concept of a "continuant". We repeat some well known facts (see for example the textbook by Hlawka and Schoissengeier [32, p. 24]). Let $a_0, \ldots, a_n$ and $c_0, c_1, \ldots, c_n$ be real numbers, where $c_0 = 1$. Choose $p_{-2}, \ldots, p_n$ and $q_{-2}, \ldots, q_n$ such that

$$\begin{pmatrix} p_{-1} & p_{-2} \\ q_{-1} & q_{-2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$p_k = a_k p_{k-1} + c_k p_{k-2}$$

$$q_k = a_k q_{k-1} + c_k p_{k-2}$$

for $k \geq 0$. Then it is easy to see that

$$\begin{pmatrix} p_k & p_{k-1} \\ q_k & q_{k-1} \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ c_0 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_k & 1 \\ c_k & 0 \end{pmatrix}$$

for all $k \geq -1$. The values p_k can be expressed as the determinant of a tridiagonal matrix, a so-called *continuant*:

$$p_k = \det \begin{pmatrix} a_0 & -1 & & & \\ c_1 & a_1 & -1 & & \\ & c_2 & \ddots & \ddots & \\ & & \ddots & \ddots & -1 \\ & & & c_k & a_k \end{pmatrix}$$

(see Perron [50, p. 11]) We want to write the values $z(t)$ as an expression of a similar form. To do this, assume that t is odd and that $t = \left(11^{k_0}0^{k_1} \cdots 1^{k_{r-2}}0^{k_{r-1}}1^{k_r}\right)_2$, where $r \geq 0$ is even, $k_0 \geq 0$ and $k_i > 0$ for $i \geq 1$. For simplicity, we assume that $z(1) = 1$. We start with the representation

$$z(t) = \begin{pmatrix} 1 & 0 \end{pmatrix} B^{k_r} A^{k_{r-1}} \cdots B^{k_2} A^{k_1} B^{k_0} \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

where $A = \left(\begin{smallmatrix} 1 & 0 \\ \alpha & \beta \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} \alpha & \beta \\ 0 & 1 \end{smallmatrix}\right)$ (compare (2.24)). After each power of B we insert the identity matrix to obtain

$$z(t) = \begin{pmatrix} 1 & 0 \end{pmatrix} B^{k_r} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} A^{k_{r-1}} \cdots B^{k_0} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Transposing this identity and performing the swaps of rows and columns given by the antidiagonal unit matrices, we obtain

$$z(t) = \begin{pmatrix}1 & 1\end{pmatrix} B(k_0)A(k_1)\cdots B(k_{r-2})A(k_{r-1})B(k_r) \begin{pmatrix}1\\0\end{pmatrix}, \tag{2.27}$$

where

$$A(k) = \begin{pmatrix}\alpha \sum_{i<k} \beta^i & 1\\ \beta^k & 0\end{pmatrix}$$

$$B(k) = \begin{pmatrix}\beta \sum_{i<k} \alpha^i & 1\\ \alpha^k & 0\end{pmatrix}.$$

In order to get a representation of $z(t)$ as a determinant, we define values a_j and c_j for $0 \le j \le k$ by $a_i = \begin{cases}\beta \sum_{i<k_j} \alpha^i, & j \text{ is even}\\ \alpha \sum_{i<k_j} \beta^i, & j \text{ is odd}\end{cases}$ and $c_i = \begin{cases}\alpha^{k_j}, & j \text{ is even}\\ \beta^{k_j}, & j \text{ is odd}\end{cases}$. we define $p_{-2} = 1$, $p_{-1} = 1$ and $p_j = a_k p_{j-1} + c_k p_{j-2}$ for $0 \le j \le r$. Then by (2.27) we have $p_r = z(t)$. We follow Perron [50, p. 10]. From the definition of $p_{-2}, \ldots, p_r$ we get the following system of linear equations in the variables $p_0, p_1, \ldots, p_r$:

$$\begin{array}{ccccccccc}
-p_0 & & & & & & & = & -a_0 - c_0\\
a_1 p_0 & - & p_1 & & & & & = & -c_1\\
c_2 p_0 & + & a_2 p_1 & - & p_2 & & & = & 0\\
& & \cdots & & \cdots & & \cdots & & \cdots\\
& & c_r p_{r-2} & + & a_r p_{r-1} & - & p_r & = & 0
\end{array}$$

By Cramer's rule we get

$$z(t) = p_r = (-1)^{r+1} \det \begin{pmatrix}
-1 & & & & & -a_0 - c_0\\
a_1 & -1 & & & & -c_1\\
c_2 & a_2 & -1 & & & 0\\
& c_3 & \ddots & \ddots & & \vdots\\
& & \ddots & \ddots & -1 & \vdots\\
& & & c_r & a_r & 0
\end{pmatrix}.$$

Rotating the columns by one place and adjusting signs, we obtain

$$z(t) = \det \begin{pmatrix}
a_0 + c_0 & -1 & & & & \\
c_1 & a_1 & -1 & & & \\
& c_2 & a_2 & -1 & & \\
& & c_3 & \ddots & \ddots & \\
& & & \ddots & \ddots & -1\\
& & & & c_r & a_r
\end{pmatrix}.$$

There is one special case of a sequence z as above that deserves special attention. For $\alpha = \beta = 1$ we obtain the so-called *Stern-Brocot-sequence*, for which the reflection property was known already to Dijkstra [20, p230ff], [19]. Different proofs are known, our second proof of the reflection property generalized the idea of proof from Urbiha [55]. The survey article [48] by Northshield lists properties of this sequence "that have most impressed the author",

among them the reflection property. The above representation of $z(t)$ as a determinant immediately yields the reflection property for the Stern-Brocot-sequence, since we only have to flip the matrix with respect to the antidiagonal, which does not change the determinant. Note however that this procedure does not work when $c_i \neq 1$, which might happen as soon as $\alpha \neq 1$ or $\beta \neq 1$. Unfortunately we could not find another proof of the reflection property using the continuant representation.

As a concluding remark on the Stern-Brocot sequence, we also note the following connection to continued fractions that can be seen from the representation (2.27): Let $t = (1^{k_0}0^{k_1}\cdots 1^{k_{r-2}}0^{k_{r-1}}1^{k_r})_2$. Then $z(t)$ is the numerator of the continued fraction $[k_0; k_1, \ldots, k_r]$. The reflection property in this case therefore is the same as the well-known statement that $[k_0; k_1, \ldots, k_r]$ and $[k_r; k_{r-1}, \ldots, k_0]$ have the same numerator.

The square mean of the correlation function for the Thue-Morse sequence

Let $q = 2$, $\vartheta = 1/2$, $\alpha = \mathrm{e}(\vartheta)/2$ and $\beta = \mathrm{e}(-\vartheta)/2$. We consider the corresponding correlation function γ_t and the sum

$$S_1(k) = \sum_{2^k \le t < 2^{k+1}} |\gamma_t|^2 .$$

We define

$$S_2(k) = \sum_{2^k \le t < 2^{k+1}} \gamma_t \overline{\gamma_{t+1}}$$

and

$$S_3(k) = \sum_{2^k \le t < 2^{k+1}} \overline{\gamma_t} \gamma_{t+1}$$

For $k \ge 1$ we have

$$\begin{aligned} S_1(k) &= \sum_{2^{k-1} \le t < 2^k} |\gamma_{2t}|^2 + \sum_{2^{k-1} \le t < 2^k} |\gamma_{2t+1}|^2 = S_1(k-1) + \sum_{2^{k-1} \le t < 2^k} |\alpha\gamma_t + \beta\gamma_{t+1}|^2 \\ &= S_1(k-1) + |\alpha|^2 \sum_{2^{k-1} \le t < 2^k} |\gamma_t|^2 + |\beta|^2 \sum_{2^{k-1} \le t < 2^k} |\gamma_{t+1}|^2 + \alpha\overline{\beta} S_2(k-1) + \overline{\alpha}\beta S_3(k-1) \\ &= (1 + |\alpha|^2 + |\beta|^2) S_1(k-1) + \alpha\overline{\beta} S_2(k-1) + \overline{\alpha}\beta S_3(k-1) \end{aligned}$$

$$\begin{aligned} S_2(k) &= \sum_{2^k \le t < 2^{k+1}} \gamma_t \overline{\gamma_{t+1}} = \sum_{2^{k-1} \le t < 2^k} \gamma_{2t} \overline{\gamma_{2t+1}} + \sum_{2^{k-1} \le t < 2^k} \gamma_{2t+1} \overline{\gamma_{2t+2}} \\ &= \sum_{2^{k-1} \le t < 2^k} \gamma_t \overline{\alpha\gamma_t + \beta\gamma_{t+1}} + \sum_{2^{k-1} \le t < 2^k} (\alpha\gamma_t + \beta\gamma_{t+1}) \overline{\gamma_{t+1}} \\ &= \overline{\alpha} S_1(k-1) + \overline{\beta} S_2(k-1) + \alpha S_2(k-1) + \beta S_1(k-1) + \beta\gamma_{2^k} - \beta\gamma_{2^{k-1}} \\ &= (\overline{\alpha} + \beta) S_1(k-1) + (\alpha + \overline{\beta}) S_2(k-1). \end{aligned}$$

and

$$S_3(k) = (\alpha + \overline{\beta}) S_1(k-1) + (\overline{\alpha} + \beta) S_3(k-1).$$

Written as a matrix-vector product, we get

$$\begin{pmatrix} S_1(k) \\ S_2(k) \\ S_3(k) \end{pmatrix} = \begin{pmatrix} 1+|\alpha|^2+|\beta|^2 & \alpha\overline{\beta} & \overline{\alpha}\beta \\ \overline{\alpha}+\beta & \alpha+\overline{\beta} & 0 \\ \alpha+\overline{\beta} & 0 & \overline{\alpha}+\beta \end{pmatrix} \begin{pmatrix} S_1(k-1) \\ S_2(k-1) \\ S_3(k-1) \end{pmatrix}$$

In the case of the Thue-Morse sequence, that is, $\alpha = \beta = -1/2$, we have $S_2(k) = S_3(k)$ and we get the simplified representation

$$\begin{pmatrix} S_1(k) \\ S_2(k) \end{pmatrix} = \begin{pmatrix} \frac{3}{2} & \frac{1}{2} \\ -1 & -1 \end{pmatrix} \begin{pmatrix} S_1(k-1) \\ S_2(k-1) \end{pmatrix}$$

Its characteristic polynomial is $X^2 - 1/2X - 1$. By some calculation involving eigenvalues and eigenvectors we get the exact representation

$$\sum_{2^k \le t < 2^{k+1}} |\gamma_t|^2 = \frac{\left(a_1\lambda_1^k + a_2\lambda_2^k\right)}{9\sqrt{17}},$$

where $\lambda_{1,2} = \frac{1\pm\sqrt{17}}{4}$ and $a_{1,2} = \frac{7\pm\sqrt{17}}{2}$.

It would also be interesting to access other summation limits than powers of two. For this more general case we expect a result of the type obtained in [22], where problems related to asymptotics related to the sum-of-digits are studied. That is, we expect

$$\sum_{t<N} |\gamma_t|^2 = N^\rho F(\log N/\log 2) + \text{error term},$$

where F is 1-periodic and continuous and $\rho = \log\frac{1+\sqrt{17}}{2}/\log 2$.

The discrete Fourier transform of the correlation function

In this short section, we study the discrete Fourier transform of γ_u, where u runs through a dyadic interval and where γ is the correlation of the function $n \mapsto \mathrm{e}(\vartheta s_2(n))$. For simplicity we restrict ourselves to the case that $q = 2$, although some results may remain valid for the more general case. We define the discrete Fourier coefficients

$$a_{h,\lambda}(\vartheta) = 2^{-\lambda} \sum_{2^\lambda \le u < 2^{\lambda+1}} \mathrm{e}\left(-hu2^{-\lambda}\right) \gamma_u(\vartheta).$$

By splitting into even and odd indices and using the recurrence relation for $\gamma_t(\vartheta)$ we get for $\lambda \geq 1$

$$\begin{aligned}
a_{h,\lambda}(\vartheta) &= 2^{-\lambda} \sum_{2^{\lambda-1} \leq u < 2^\lambda} \mathrm{e}\left(-hu2^{\lambda-1}\right) \gamma_u(\vartheta) \\
&\quad + \mathrm{e}\left(-h2^{-\lambda}\right) 2^{-\lambda} \sum_{2^{\lambda-1} \leq u < 2^\lambda} \mathrm{e}\left(-huq^{\lambda-1}\right) \left(\frac{1}{2}\,\mathrm{e}(\vartheta)\gamma_u(\vartheta) + \frac{1}{2}\,\mathrm{e}(-\vartheta)\gamma_{u+1}(\vartheta)\right) \\
&= \frac{1}{2} a_{h,\lambda-1}(\vartheta) + \frac{1}{4}\,\mathrm{e}\left(\vartheta - h2^{-\lambda}\right) a_{h,\lambda-1}(\vartheta) \\
&\quad + \frac{1}{4}\,\mathrm{e}(-\vartheta - h2^{-\lambda}) 2^{-\lambda-1} \sum_{2^{\lambda-1} < u \leq 2^\lambda} \mathrm{e}(-h(u-1)2^{-\lambda-1}) \gamma_u(\vartheta) \\
&= \frac{1}{2}\left(1 + \cos(2\pi(\vartheta - h2^{-\lambda}))\right) a_{h,\lambda-1}(\vartheta).
\end{aligned}$$

Since $1 + \cos(x) = 2\cos^2(x/2)$ we get by induction

$$a_{h,\lambda}(\vartheta) = \gamma_1(\vartheta) \prod_{1 \leq r \leq \lambda} \cos^2(\pi(\vartheta - h2^{-r}))$$

for $\lambda \geq 0$ and therefore by the inverse Fourier transform and the shift $h \mapsto h - 2^\lambda$

$$\gamma_t(\vartheta) = \gamma_1(\vartheta) \sum_{h < 2^\lambda} \mathrm{e}\left(ht2^{-\lambda}\right) \prod_{1 \leq r \leq \lambda} \cos^2\left(\pi\left(\vartheta - h2^{-r}\right)\right) \tag{2.28}$$

for $\lambda \geq 0$ and $2^\lambda \leq t < 2^{\lambda+1}$. Note that this is a representation of the correlation $\gamma_t(\vartheta)$ in terms of trigonometric functions only, in particular it does not involve the sum-of-digits function.

A product representation for a generating function

As another small remark, we derive an explicit representation of the generating function for a sequence $(z_t)_{t \geq 1}$ satisfying $z_{2t} = z_t$ and $z_{2t+1} = \alpha z_t + \beta z_{t+1}$ for some $\alpha, \beta \in \mathbb{C}$ such that $|\alpha| \leq 1$ and $|\beta| \leq 1$, at least for the case that $\beta \notin \mathbb{T} \setminus \{1\}$. The sequences $\gamma_t(\vartheta)$ from above are special examples of such sequences.

We define a power series

$$G(x) = \sum_{t \geq 1} z_t x^{t-1}.$$

This defines a function that is holomorphic in some circle $\{z : |z| < R\}$ around the origin.

For $0 < |x| < R$ we get

$$\begin{aligned} G(x) &= \sum_{t\geq 1} z_t x^{t-1} \\ &= \sum_{t\geq 1} z_{2t} x^{2t-1} + \sum_{t\geq 0} z_{2t+1} x^{2t} \\ &= x \sum_{t\geq 1} z_t x^{2(t-1)} + z_1 + \sum_{t\geq 1} (\alpha z_t + \beta z_{t+1})\, x^{2t} \\ &= xG(x^2) + z_1 + \alpha x^2 G(x^2) + \beta \sum_{t\geq 2} z_t x^{2t-2} \\ &= (x + \alpha x^2 + \beta) G(x^2) + z_1(1-\beta). \end{aligned}$$

We write for a moment $f(x) = x + \alpha x^2 + \beta$ and $g(x) = z_1(1-\beta)$. Then for all $\lambda > 0$ we have

$$\begin{aligned} G(x) = g(x) + f(x)G(x^2) = g(x) + f(x)g(x^2) + f(x)f(x^2)G(x^4) = \cdots \\ = g(x) \sum_{\mu<\lambda} \prod_{i<\mu} f\left(x^{2^i}\right) + \prod_{i<\lambda} f\left(x^{2^i}\right) G\left(x^{2^\lambda}\right). \end{aligned}$$

Case 1. Assume first that $|\beta| < 1$. We have $G(z) \ll 1$ as $z \to 0$, therefore the second term is

$$\ll \prod_{i<\lambda} \left(x^{2^i} + \alpha x^{2^{i+1}} + \beta\right).$$

For each x such that $|x| < 1$ and for $|\beta| < 1$ this expression converges to zero as $\lambda \to \infty$. We obtain the following representation of the generating function as an infinite sum of polynomials.

$$G(x) = z_1(1-\beta) \sum_{\lambda\geq 0} \prod_{i<\lambda} \left(x^{2^i} + \alpha x^{2^{i+1}} + \beta\right).$$

Case 2. $\beta = 1$. In this case all summands but the last are zero. Moreover, for each $j \geq 1$ there is a λ such that $[x^j]\, G\left(x^{2^i}\right) = 0$ for all $i \geq j$. For the zeroth coefficient we have $[x^0]\, G\left(x^{2^i}\right) = z_1$. It follows that we obtain the following representation of G as an infinite product of polynomials.

$$G(x) = z_1 \prod_{i\geq 0} \left(x^{2^i} + \alpha x^{2^{i+1}} + 1\right). \tag{2.29}$$

(The special case $\alpha = 1$ and $z_1 = 1$ of this formula, corresponding to the Stern-Brocot sequence, is also contained in [19]).

Case 3. $|\beta| = 1$, $\beta \neq 1$. In this case the product is divergent, so that we do not get a useful representation as an infinite series or product easily.

We check some coefficients of this representation. We have $[x^0]\, G(x) = z_1(1-\beta) \sum_{\lambda\geq 0} \beta^\lambda = z_1$. For the second coefficient we have to choose x^{2^0} as the first factor in the product

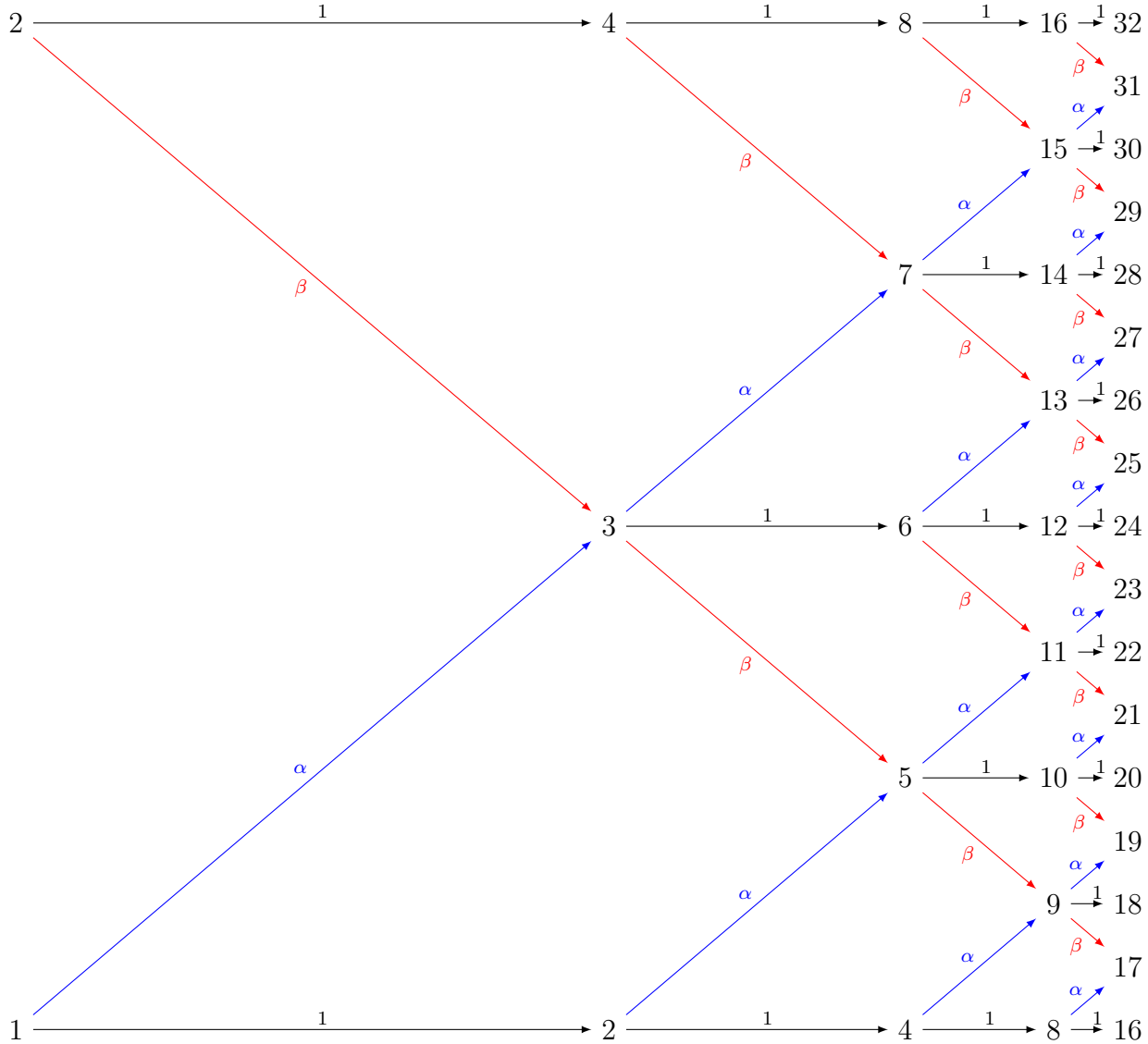
and β as the remaining factors. For $\lambda = 0$ the contribution of the product is zero. We obtain $[x^1]\,G(x) = z_1(1-\beta)\sum_{\lambda\geq 1}\beta^{\lambda-1} = z_1$. In the case of the third coefficient, the product contributes 0 for $\lambda = 0$, α for $\lambda = 1$ and $\beta^{\lambda-1}(1+\alpha)$ for $\lambda \geq 2$. We obtain $[x^2]\,G(x) = z_1(1-\beta)\left(\alpha + (1+\alpha)\sum_{\lambda\geq 2}\beta^{\lambda-1}\right) = (\alpha+\beta)z_1$. On can easily imagine that this method of determining the values z_t gets more and more involved for higher and higher t.

In the (sum-of-digits-) case that $\alpha = \frac{1}{2}\,\mathrm{e}(\vartheta)$, $\beta = \frac{1}{2}\,\mathrm{e}(-\vartheta)$ and $z_1 = \frac{\mathrm{e}(\vartheta)}{2-\mathrm{e}\,-\vartheta} = \frac{\alpha}{1-\beta}$ we have $x^{2^i} + \alpha x^{2^{i+1}} + \beta = \alpha\left(x^{2^i} + 2\beta\right)^2$ and therefore

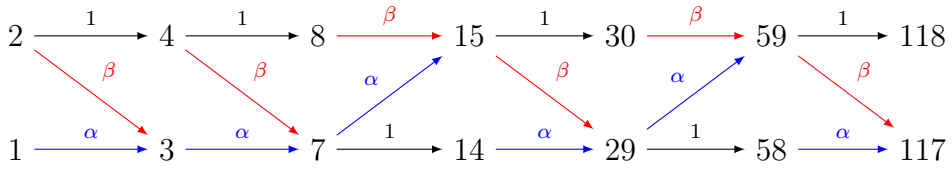
$$G(x) = \sum_{\lambda\geq 0}\alpha^{\lambda+1}\prod_{i<\lambda}\left(x^{2^i}+2\beta\right)^2. \tag{2.30}$$

A graphical representation of the recurrence relation $z_{2t} = z_t$, $z_{2t+1} = \alpha z_t + \beta z_{t+1}$

We finish this section with a graphical representation of the recurrence relation for z, which is self-explanatory (see also [48]).



Using this picture, we can for each positive integer t determine the weighted sum over all paths, leading to it (starting at the top row), which yields the value z_t when multiplied with z_1. For example, z_{117} can be easily determined using the following subgraph, noting that $117 = (1110101)_2$.



Chapter 3

The sum of digits of n and $n+t$

In the previous chapter we were concerned with the correlations

$$\gamma_t(\vartheta) = \lim_{N\to\infty} \mathrm{e}\left(\vartheta(s_q(n+t) - s_q(n))\right)$$

for the sum-of-digits function, which contains information on the relation of the sum of digits of n and $n+t$ to each other. (See Lemma 1.24 for a proof establishing the existence of the limit.) In the present chapter we continue to study this relation. We are concerned with the question due to T.W. Cusick [14] concerning the values c_t defined in (2.4): let c_t be the asymptotic density of the set of nonnegative integers n such that $s_2(n+t) \geq s_2(n)$. Is it true that

$$c_t > 1/2 \tag{3.1}$$

for all integers $t \geq 0$? (Note that the limit defining c_t always exists by Lemma (2.6).) We prove that (3.1) holds for t in a set of asymptotic density 1, using an averaging argument and Chebyshev's inequality.

3.1 Introduction

For a nonnegative integer n let $s(n)$ be the number of ones in the binary representation of n. The number $s(n)$ is the *sum of digits* of n in base 2.

In chapter 2, equation (2.3) we derived the expression

$$s_q(n+t) - s_q(n) = s_q(t) - (q-1)\max\left\{k : q^k \mid \binom{n+t}{t}\right\},$$

which clarifies the well-known connection between sum-of-digits functions and Pascal's triangle. In the present chapter, we will only be concerned with the sum of digits in base 2. We specialize therefore to the case $q = 2$ and obtain

$$s(n+t) - s(n) = s(t) - \max\left\{k : 2^k \mid \binom{n+t}{t}\right\}. \tag{3.2}$$

The left hand side is negative if and only if $\max\left\{k : 2^k \mid \binom{n+t}{t}\right\} > s_2(t)$, that is, if and only if $2^{s(t)+1} \mid \binom{n+t}{n}$. We obtain therefore

$$\#\left\{n < N : s(n+t) \geq s(n)\right\} = N - \#\left\{n < N : 2^{s(t)+1} \mid \binom{n+t}{t}\right\}. \tag{3.3}$$

The main purpose of this chapter is to shed some light on the following question posed by T. W. Cusick:

Question 1. *Assume that t is a non-negative integer and define*

$$c_t = \lim_{N\to\infty} \frac{1}{N}\#\{n < N : s(n+t) \geq s(n)\}.$$

Is it true that $c_t > 1/2$ for all $t \geq 1$?

By equation (3.3) this problem can be rephrased in terms of the number of zeroes in a column of Pascal's triangle modulo 2^k. Problems of this kind have received some attention in the literature (see Barat and Grabner [3] and the references contained therein) and have proven to be difficult to handle. In this chapter, we present some partial results regarding Cusick's question.

The following lemma is an adaptation of Lemma 2.6 to the case of the binary sum of digits. It establishes the fundamental recurrence relation that we will use throughout this chapter.

Lemma 3.1. *Let $t \geq 0$ be an integer. There exists a partition $\mathcal{N}_t$ of the set of nonnegative integers having the properties that*

- *Each $N \in \mathcal{N}_t$ is of the form $a + 2^k\mathbb{N}$, where $0 \leq a < 2^k$ and $k \geq 0$.*
- *For all integers k the set*

$$A(k,t) = \{n \in \mathbb{N} : s(n+t) - s(n) = k\}$$

is a finite (possibly empty) union of elements of sets from the partition $\mathcal{N}_t$.

In particular, each of the sets $A(k,t)$ possesses an asymptotic density $\delta(k,t)$. Moreover, for all $k, t \geq 0$ the densities satisfy the following recurrence relation:

$$\begin{array}{lcl} \delta(k,0) & = & \delta_{k,0}, \\ \delta(k,2t) & = & \delta(k,t), \\ \delta(k,2t+1) & = & \frac{1}{2}\delta(k-1,t) + \frac{1}{2}\delta(k+1,t+1). \end{array} \tag{3.4}$$

We also adapt the proof of Lemma 2.6 to the case that $q = 2$, which is mathematically not necessary but makes the chapter more self contained.[1]

Proof. We set $d(n,t) = s(n+t) - s(n)$. We have $d(n,0) = 0$ for all n, therefore $A(k,0) = \mathbb{N}$ if $k = 0$ and $A(k,0) = \emptyset$ otherwise, which implies the first line of (3.4). For all $n, t \geq 0$ the values of d satisfy the property that

$$\begin{array}{lcl} d(2n,2t) & = & d(n,t) \\ d(2n+1,2t) & = & d(n,t) \\ d(2n,2t+1) & = & d(n,t)+1 \\ d(2n+1,2t+1) & = & d(n,t+1)-1 \end{array} \tag{3.5}$$

[1] At least a closer look at this special case was very helpful for the author.

which follows easily from the elementary property $s(2m+r) = s(m) + r$ for $r \in \{0,1\}$.

Moreover, using the binary representation of n in base 2, we obtain

$$d(n,1) = 1 - \max\{k \geq 0 : 2^k \mid n+1\}. \tag{3.6}$$

We prove the statements by induction on t. In the case $t = 1$ equation (3.6) implies

$$A(1-r,t) = \left\{ \begin{array}{cc} -1 + 2^r + 2^{r+1}\mathbb{N} & r \geq 0 \\ \emptyset & \text{otherwise} \end{array} \right\}$$

since the set of nonnegative n exactly divisible by 2^r is $2^r + 2^{r+1}\mathbb{N}$.

Let $t > 1$ be even, $t = 2u$, and $k \in \mathbb{Z}$. Then

$$\begin{aligned} A(k,2u) &= \{n : d(n,2u) = k\} \\ &= 2\{n : d(2n,2u) = k\} \cup (2\{n : d(2n+1,2u) = k\} + 1) \\ &= 2\{n : d(n,u) = k\} \cup (2\{n : d(n,u) = k\} + 1), \end{aligned} \tag{3.7}$$

which is by the induction hypothesis a finite union of arithmetic progressions of the form $a + 2^k\mathbb{N}$. If t is odd, $t = 2u+1$, we get by analogous reasoning

$$\begin{aligned} A(k,2u+1) &= \{n : d(n,2u+1) = k\} \\ &= 2\{n : d(n,u) = k-1\} \cup (2\{n : d(n,u+1) = k+1\} + 1). \end{aligned} \tag{3.8}$$

The unions in (3.7) and (3.8) respectively are disjoint, therefore the statement on the densities follows. This finishes the proof. □

The recurrence for the densities $\delta(k,t)$ allow us to compute their values for any given value of t. We list some entries of the double family δ.

$k \setminus t$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	$\frac{1}{16}$	0
3	0	0	0	0	0	0	0	$\frac{1}{8}$	0	0	0	$\frac{1}{8}$	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{32}$	0
2	0	0	0	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{16}$	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{5}{64}$	0
1	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{5}{32}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{5}{32}$	$\frac{21}{128}$	$\frac{1}{2}$
0	1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{5}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{5}{16}$	$\frac{21}{64}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{8}$	$\frac{5}{32}$	$\frac{5}{16}$	$\frac{5}{32}$	$\frac{21}{64}$	$\frac{85}{256}$	$\frac{1}{4}$
-1	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{5}{32}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{5}{32}$	$\frac{21}{128}$	$\frac{1}{8}$	$\frac{3}{32}$	$\frac{3}{16}$	$\frac{13}{64}$	$\frac{5}{32}$	$\frac{13}{64}$	$\frac{21}{128}$	$\frac{85}{512}$	$\frac{1}{8}$
-2	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{5}{64}$	$\frac{1}{16}$	$\frac{3}{32}$	$\frac{5}{64}$	$\frac{21}{256}$	$\frac{1}{16}$	$\frac{7}{64}$	$\frac{3}{32}$	$\frac{13}{128}$	$\frac{5}{64}$	$\frac{13}{128}$	$\frac{21}{256}$	$\frac{85}{1024}$	$\frac{1}{16}$
-3	0	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{5}{128}$	$\frac{1}{32}$	$\frac{3}{64}$	$\frac{5}{128}$	$\frac{21}{512}$	$\frac{1}{32}$	$\frac{7}{128}$	$\frac{3}{64}$	$\frac{13}{256}$	$\frac{5}{128}$	$\frac{13}{256}$	$\frac{21}{512}$	$\frac{85}{2048}$	$\frac{1}{32}$

By (2.3) we have the inequality $s(n+t)-s(n) \le s(t)$, therefore the value c_t can be obtained by summing finitely many values of $\delta(k,t)$:

$$c_t = \sum_{k\ge 0} \delta(k,t).$$

The sequence $(c_t)_{t\ge 0}$ therefore begins as follows:

$$\left(1, \frac{3}{4}, \frac{3}{4}, \frac{11}{16}, \frac{3}{4}, \frac{5}{8}, \frac{11}{16}, \frac{43}{64}, \frac{3}{4}, \frac{11}{16}, \frac{5}{8}, \frac{19}{32}, \frac{11}{16}, \frac{19}{32}, \dots\right).$$

A numerical experiment conducted by the author, using this method of computing c_t, reveals that the answer to Question 1 is positive for all $t < 2^{30}$. The minimal value of c_t for t in this range is attained at $t = 1039104991 = (111101111011110111101111011111)_2$ and also (which is then guaranteed by (2.5)) at the integer obtained from this one by reversing the binary digits. The value of c_t at these positions is $18169025645289/2^{45} = 0.516394\ldots$.

Examining the family δ, we also see that

$$\tilde{c}_t := \sum_{k\ge 1} \delta(k,t)$$

always seems to be $\le 1/2$. At least this is suggested by the first 2^{30} values, which we computed in the same way as c_t. This leads us to the following reasonable question:

Question 2. *Assume that t is a nonnegative integer and define*

$$\tilde{c}_t = \lim_{N\to\infty} \frac{1}{N} \#\{n < N : s(n+t) > s(n)\}.$$

Do we have $\tilde{c}_t \le 1/2$ for all $t \ge 1$?

Translating this to a statement on divisibility in Pascal's triangle, using (3.2), we obtain the following refined form of the problem, which asks both questions simultaneously:

Question 3. *Let $t \ge 0$ be an integer. What is the largest exponent k such that*

$$\operatorname{dens}\left\{n : 2^k \mid \binom{n+t}{t}\right\} \ge 1/2$$

holds?

The answer to this question is $s(t)$ if and only if the answer to both Question 1 and Question 2 is positive. This question is particularly intriguing, since the sum-of-digits function does not feature in it. It is a question on divisibility by powers of two in a column of Pascal's triangle and it asks for the "right" power of two to take as a modulus. Nevertheless the sum-of-digits function should feature in the answer.

Note that the question is well-posed, that is, there is an exponent k such that the densities corresponding to k and $k+1$ are $\ge 1/2$ and $< 1/2$ respectively. This can be seen easily from the recurrence relation governing $\delta(k,t)$.

In this chapter we show several partial results concerning Question 1. The first result is a nontrivial lower bound for c_t for "half of" the positive integers t. Next, we prove $c_{t_j} > 1/2$ for a certain subsequence $(t_j)_j$ of the integers, given by integers having the binary expansion $((10)^j)_2$. The main result establishes $c_t > 1/2$ for almost all positive integers t.

3.2 Results

We only consider partial answers to Question 1. It is certainly possible to obtain analogous results for Question 2. However, since they do not give more insight into the problem we focus on Question 1.

It seems difficult to obtain a nontrivial bound for individual values c_t. However, considering two values c_t and $c_{t'}$ simultaneously, we can already prove a nontrivial result.

Theorem 3.2. *Assume that t be a positive integer and let $t' = 3 \cdot 2^\lambda - t$, where λ is chosen in such a way that $2^\lambda \le t < 2^{\lambda+1}$. Then*

$$c_t > \frac{15}{32} \quad \text{or} \quad c_{t'} > \frac{15}{32}.$$

It turns out that appropriate bi- and trivariate generating functions are a useful tool to attack Question 1. As a first application of this method we prove the following result.

Theorem 3.3. *Let $j \ge 0$ and t_j be defined by the binary expansion $t_j = ((01)^j)_2$. There exists a $j_0 \ge 1$ such that for all $j \ge j_0$ we have*

$$c_{t_j} > \frac{1}{2}.$$

However, the main result of this chapter is the following asymptotic statement.

Theorem 3.4. *We have, as $T \to \infty$,*

$$T - |\{t \le T : c_t > 1/2\}| = O\left(\frac{T}{\sqrt{\log T}}\right),$$

that is, $c_t > 1/2$ holds for t in a subset of $\mathbb{N}$ of asymptotic density 1.

The proof is based on an adequate averaging argument. More precisely we study the distribution of c_t for $2^\lambda \le t < 2^{\lambda+1}$ and show, using Chebyshev's inequality, that the values of c_t concentrate around $1/2 + 1/(2\sqrt{\lambda\pi})$. Whereas the average value is relatively easy to determine, the computation of the variance, which we use in order to obtain the concentration result, relies on a quite involved analysis of a trivariate generating function.[2]

3.3 Proof of Theorem 3.2

In order to prove Theorem 3.2, we define a simplified array as follows. Let $\varphi(k,1) = \delta_{k,0}$, $\varphi(k,2t) = \varphi(k,t)$ and $\varphi(k,2t+1) = \frac{1}{2}\varphi(k-1,t) + \frac{1}{2}\varphi(k+1,t+1)$ for $t \ge 1$. We list some

[2]The author is very grateful to Michael Drmota for his advice concerning the proof of Proposition 3.7, and to Manuel Kauers, who derived this result with the help of a computer algebra system.

values of φ in a table and omit zeroes for more clarity.

$k \setminus t$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
4																					
3															$\frac{1}{8}$						
2							$\frac{1}{4}$				$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$					$\frac{1}{4}$		$\frac{1}{4}$
1			$\frac{1}{2}$		$\frac{1}{2}$	$\frac{1}{2}$			$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$		$\frac{1}{8}$		$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{4}$
0	1	1		1	$\frac{1}{4}$		$\frac{1}{4}$	1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{5}{16}$	$\frac{1}{4}$	$\frac{1}{16}$
-1			$\frac{1}{2}$			$\frac{1}{2}$	$\frac{1}{2}$		$\frac{1}{8}$		$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{2}$		$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$		$\frac{1}{4}$
-2					$\frac{1}{4}$					$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$				$\frac{1}{16}$		$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{16}$
-3									$\frac{1}{8}$									$\frac{1}{8}$	$\frac{1}{8}$		$\frac{1}{8}$
-4																	$\frac{1}{16}$				
-5																					

If λ is chosen such that $2^\lambda \le t < 2^{\lambda+1}$, we define $t' = 3 \cdot 2^\lambda - t$. The double family φ has the properties that

$$\sum_{k\in\mathbb{Z}} \varphi(k,t) = 1,$$

$$\delta(k,t) = \sum_{\ell+s=k} \varphi(\ell,t)\delta(s,1)$$

and

$$\varphi(k,t) = \varphi(-k,t').$$

The first two properties are easy to prove by induction. We prove the last statement by induction on t. If t is even, $t = 2s$, the statement follows since $(2s)' = 2s'$. In the other case we write $t = 2s+1$. Then

$$s' = \left(\frac{t-1}{2}\right)' = 3\cdot 2^{\lambda-1} - \frac{t-1}{2} = \frac{t'+1}{2}$$

and $(s+1)' = \frac{t'-1}{2}$.

From the symmetry statement for the smaller values we obtain

$$\begin{aligned}
\varphi(k,t) &= \tfrac{1}{2}\varphi(k-1,s) + \tfrac{1}{2}\varphi(k+1,s+1) \\
&= \tfrac{1}{2}\varphi(-k+1,s') + \tfrac{1}{2}\varphi(-k-1,(s+1)') \\
&= \tfrac{1}{2}\varphi\left(-k+1,\tfrac{t'+1}{2}\right) + \tfrac{1}{2}\varphi\left(-k-1,\tfrac{t'-1}{2}\right) \\
&= \varphi(-k,t').
\end{aligned}$$

Using the second and the third property, we calculate

$$c_t + c_{t'} = \sum_{\ell \geq -1} (\varphi(\ell, t) + \varphi(-\ell, t)) \left(1 - 2^{-\ell-2}\right) = \frac{3}{2}\varphi(0,t) + \frac{11}{8}\varphi(1,t) + \frac{11}{8}\varphi(-1,t) + \sum_{\ell \geq 2} \left(1 - 2^{-\ell-2}\right) \varphi(\ell, t) + \sum_{\ell \geq 2} \left(1 - 2^{-\ell-2}\right) \varphi(-\ell, t).$$

Since at least one of $\varphi(-1,t)$, $\varphi(0,t)$ and $\varphi(1,t)$ is nonzero, which can be seen by an easy induction, the sum of the first three summands is $> 15/16\,(\varphi(-1,t) + \varphi(0,t) + \varphi(1,t))$. Moreover, the smallest coefficient appearing in the sum is $15/16$, which corresponds to $\ell = 2$. Therefore $c_t + c_{t'} > \frac{15}{16} \sum_{k \in \mathbb{Z}} \varphi(k,t) = \frac{15}{16}$ and the theorem is proved.

3.4 Proof of Theorem 3.3

3.4.1 A generating function for c_t for special values of t

By the property $c_{2t} = c_t$ the sequence $t = (2^j)_j$ has the property that $c_{t_j} > 1/2$. In this section we exhibit a more interesting example of such a sequence, at least one satisfying $c_{t_j} > 1/2$ for j large enough. As it turns out, the sequence t we are going to define even has the property that $c_{t_j} \to 1/2$ from above, and we give a more precise asymptotic estimate of these values. The existence of such a sequence is not guaranteed by the result on sums over $2^\lambda \leq t < 2^{\lambda+1}$.

For $j \geq 0$ we define integers t_j and u_j (the latter being auxiliary values) by

$$t_j = \left((01)^j\right)_2 \quad \text{and} \quad u_j = \left((01)^j 1\right)_2.$$

Note that $t_0 = 0$ and $u_0 = 1$. From the recurrence relation for δ we get for $j \geq 1$ the relations

$$\delta(k, t_j) = \frac{1}{2}\delta(k-1, t_{j-1}) + \frac{1}{2}\delta(k+1, u_{j-1}) \tag{3.9}$$

and

$$\delta(k, u_j) = \frac{1}{2}\delta(k-1, t_j) + \frac{1}{2}\delta(k+1, u_{j-1}). \tag{3.10}$$

We introduce the bivariate generating functions

$$F(x,y) = \sum_{j \geq 0, k \geq 0} x^j y^k \delta(j-k, t_j)$$

and

$$G(x,y) = \sum_{j \geq 0, k \geq 0} x^j y^k \delta(j+1-k, u_j).$$

We will derive a representation of F as a rational function.

For brevity, we set

$$A = \sum_{k \geq 0} y^k \delta(1-k, 1) = \frac{1}{2-y}.$$

The relations from above carry over to identities for the generating functions F and G as follows. We split the summation over j at $j=2$ and apply the recurrence relations (3.9) and (3.10) to obtain

$$\begin{aligned}
F(x,y) &= \sum_{k\geq 0} y^k \delta(-k,0) + \sum_{j\geq 1,k\geq 0} x^j y^k \delta(j-k,t_j) \\
&= 1 + \frac{1}{2}\sum_{j\geq 1,k\geq 0} x^j y^k \delta(j-k-1,t_{j-1}) + \frac{1}{2}\sum_{j\geq 1,k\geq 0} x^j y^k \delta(j-k+1,u_{j-1}) \\
&= 1 + \frac{x}{2}\sum_{j\geq 1,k\geq 0} x^{j-1} y^k \delta(j-1-k,t_{j-1}) + \frac{1}{2}\sum_{j\geq 1} x^j y^0 \delta(j+1,u_{j-1}) \\
&+ \frac{x}{2}\sum_{j\geq 1,k\geq 1} x^{j-1} y^k \delta(j-1+1-(k-1),u_{j-1}) \\
&= 1 + \frac{x}{2}\sum_{j\geq 0,k\geq 0} x^j y^k \delta(j-k,t_j) + \frac{xy}{2}\sum_{j\geq 0,k\geq 1} x^j y^{k-1} \delta(j+1-(k-1),u_j) \\
&= 1 + \frac{x}{2}F(x,y) + \frac{xy}{2}\sum_{j\geq 0,k\geq 0} x^j y^k \delta(j+1-k,u_j) \\
&= 1 + \frac{x}{2}F(x,y) + \frac{xy}{2}G(x,y).
\end{aligned}$$

The sum over $j\geq 1$ at $k=0$ equals zero, since $s_2(u_j)=j+1$ and $\delta(k,t)=0$ for $k>s_2(t)$. We obtain

$$F(x,y) = \frac{\frac{xy}{2}G(x,y)+1}{1-\frac{x}{2}} = \frac{xy}{2-x}G(x,y) + \frac{2}{2-x}.$$

Similarly, we have

$$\begin{aligned}
G(x,y) &= \sum_{k\geq 0} y^k \delta(1-k,1) + \frac{1}{2}\sum_{j\geq 1,k\geq 0} x^j y^k \delta(j-k,t_j) + \frac{1}{2}\sum_{j\geq 1,k\geq 0} x^j y^k \delta(j-k+2,u_{j-1}) \\
&= A + \frac{1}{2}F(x,y) - \frac{1}{2}\sum_{k\geq 0} x^0 y^k \delta(-k,0) + \frac{x}{2}\sum_{j\geq 0,k\geq 0} x^j y^k \delta(j+1-(k-2),u_j) \\
&= A - \frac{1}{2} + \frac{1}{2}F(x,y) + \frac{xy^2}{2}G(x,y),
\end{aligned}$$

therefore

$$G(x,y) = \left(A - \frac{1}{2} + \frac{1}{2}F(x,y)\right) / \left(1 - \frac{xy^2}{2}\right) = \frac{2A-1}{2-xy^2} + \frac{1}{2-xy^2}F(x,y).$$

We insert this in the expression for F and obtain by a short calculation

$$F(x,y) = \frac{xy}{2-x}\cdot\frac{\frac{2}{2-y}-1}{2-xy^2} + \frac{2}{2-x} + \frac{xy}{2-x}\cdot\frac{1}{2-xy^2}F(x,y) = \frac{1}{2-y}\cdot\frac{2xy^3-3xy^2-4y+8}{x^2y^2-2xy^2-2x+4}.$$

We have by construction

$$\left[x^j y^k\right] F(x,y) = \delta\left(j-k,t_j\right)$$

for $j \geq 1$ and $k \geq 0$, moreover $\delta(k, t_j) = 0$ for $k > j$, therefore

$$c_{t_j} = \sum_{0 \leq k \leq j} \delta(j-k, t_j) = \sum_{0 \leq k \leq j} \left[x^j y^k\right] F(x,y)$$
$$= \left[x^j y^j\right] \frac{1}{(1-y)(2-y)} \cdot \frac{2xy^3 - 3xy^2 - 4y + 8}{x^2y^2 - 2xy^2 - xy - 2x + 4}. \quad (3.11)$$

These coefficients define a new power series,

$$H(z) = \sum_{j \geq 0} c_{t_j} z^j,$$

which is the *diagonal* of the rational function

$$\tilde{F}(x,y) = \frac{1}{(1-y)(2-y)} \cdot \frac{2xy^3 - 3xy^2 - 4y + 8}{x^2y^2 - 2xy^2 - xy - 2x + 4}.$$

3.4.2 The diagonal generating function

It is a well known result that the diagonal generating function of a bivariate rational generating power series is algebraic. (This result appears in Furstenberg [28] and later in Hautus and Klarner [31]. For more details see Pemantle and Wilson [49, p. 41].)

A function $H(z)$ is called algebraic if there are polynomials $p_0, \ldots, p_k \in \mathbb{C}[X]$ such that

$$p_0 + p_1 H + p_2 H^2 + \cdots + p_k H^k = 0, \quad (3.12)$$

In other words, $H \in \mathbb{C}[[X]]$ is algebraic over $\mathbb{C}[X]$. In fact, it will turn out that H (in our specail case) is algebraic over $\mathbb{Q}[X]$.

In order to obtain a guess for the equation for H we made an ansatz to find $p_0, \ldots, p_k$. In order to do this, we need the first few coefficients of H, which can be obtained for example using (3.4).

Using a computer algebra system for solving our ansatz, we obtained that $H(z)$ probably satisfies the relation

$$q_2(z)H(z)^2 + q_1(z)H(z) + q_0(z) = 0,$$

where

$$\begin{aligned} q_0(z) &= -2z^3 \\ q_1(z) &= 2z^5 - 3z^4 - 8z^3 - 7z^2 + 32z - 16 \\ q_2(z) &= z^6 - 5z^5 - 3z^4 + 5z^3 + 30z^2 - 44z + 16. \end{aligned}$$

This algebraic equation eliminates at least the first 100 coefficients of H (which is much more than is needed for guessing q_0, q_1, q_2). It is not very difficult to obtain this result in a rigorous way. For this, we use the following theorem from the paper [31] by Hautus and Klarner that is referred to above.

Theorem 3.5 (Hautus and Klarner, 1971)**.** *Suppose that*

$$F(x,y) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} f(m,n) x^m y^n$$

converges for all x and y such that $|x| < A$, $|y| < B$, then for all z such that $|z| < AB$ we have

$$\frac{1}{2\pi i}\int_C F(s, z/s)\,\mathrm{d}s/s = \sum_{n=0}^{\infty} f(n,n)z^n,$$

where C is the circle $\{s : |s| = (A + |z|/B)/2\}$.

In order to apply this to our function $\tilde{F}$, we choose $A = 1$ and $B = 1/2$. The "diagonal function" H can therefore be obtained by summing, for each z such that $|z| < 1/2$, certain residues of the function

$$\tilde{F}_z : s \mapsto \frac{1}{s}F\left(s, \frac{z}{s}\right) = \frac{2z^3 - 3sz^2 - 4sz + 8s^2}{(s-z)(2s-z)(-2s^2 + (z^2 - z + 4)s - 2z^2)}.$$

These residues can only come from the zeroes of the denominator, which are

$$\begin{aligned} s &= z \\ s &= z/2 \\ s &= \tfrac{1}{4}\left(z^2 - z + 4 \pm \sqrt{z^4 - 2z^3 - 7z^2 - 8z + 16}\right). \end{aligned}$$

Define the contour C as in the theorem, $C = \{s : |s| = 1/2 + |z|\}$. Since $|z| < 1/2$, the first two zeros lie inside the contour. The third zero (taking the "+") certainly has absolute value greater than 1 and therefore lies outside. In order to treat the fourth zero, we note that $|a| \le 65/16$, where $a = z^4 - 2z^3 - 7z^2 - 8z$. Let b be chosen such that $\sqrt{16 + a} = 4 + b$. Writing $16 + a$ in polar coordinates and taking the square root, it follows that $|b| \le 1$ by some minor estimates. It follows that the fourth zero lies inside the contour. Using again a computer algebra system (although it would be possible to do this by hand), we calculated the residues of F_z at the three relevant points, which are functions of z, yielding

$$H(z) = \frac{-q_1(z) + \sqrt{q_1(z)^2 - 4q_0(z)q_2(z)}}{2q_2(z)}$$

with the polynomials q_i from above. The asymptotic behaviour of the coefficients of H can be analyzed using singularity analysis (see Flajolet and Odlyzko [23] and Flajolet and Sedgewick [24]).

We have the factorizations

$$\begin{aligned} q_1(z) &= (z^2 + 3z + 4)(z-1)(2z^2 - 7z + 4) \\ q_2(z) &= (z^2 + 3z + 4)(z-1)^2(z^2 - 6z + 4) \\ q_1(z)^2 - 4q_0(z)q_2(z) &= (z^2 + 3z + 4)(z-1)^3(z-4)(2z^2 + 3z - 4)^2, \end{aligned}$$

which gives (a kind of) partial fraction decomposition

$$\begin{aligned} H(z) = -\frac{1}{2(z-1)} - \frac{z}{2(z^2 - 6z + 4)} + \sqrt{(z-1)(z-4)(z^2 + 3z + 4)} \\ \times \left(\frac{z}{16(z^2 + 3z + 4)} + \frac{1}{12(z^2 + 3z + 4)} + \frac{1}{6(z^2 - 6z + 4)} - \frac{1}{16(z-1)}\right). \end{aligned}$$

In order to apply singularity analysis it is necessary to determine the singularities of $H(z)$. For example, $z=1$ is a polar singularity as well as a singularity which appears as $1/\sqrt{1-z}$ and as $\sqrt{1-z}$. The root $3-\sqrt{5}$ which is the (smaller) root of z^2-6z+4 is a removable polar singularity so that it does not contribute. The other singularities ($z=4$, $z=3+\sqrt{5}$, and $z=(3\pm i\sqrt{7})/2$ have modulus larger than 1 which implies that $z=1$ is the dominant singularity. The term $-1/2(z-1)$ contributes the (constant) term $1/2$, and it remains to determine the asymptotic behaviour of the coefficients of $-\sqrt{(z-1)(z-4)(z^2+3z+4)}/(16(z-1))$. In order to do this, we expand this term in the Δ-region

$$\Delta = \{z : |z| < 3/2, |\arg(z-1)| < \pi/8\},$$

in which the function $H(z)$ is analytic, as follows: we have

$$\frac{\sqrt{(z-1)(z-4)(z^2+3z+4)}}{z-1} = \frac{-c_1}{\sqrt{1-z}} + O(1)$$

as $z \to 1$, $z \in \Delta$, where

$$c_1 = \sqrt{(4-z)(z^2+3z+4)}\Big|_{z=1} = 2\sqrt{6}.$$

We apply Theorem VI.3 from [24] to the error term, moreover we use the asymptotic formula

$$[z^j](1-z)^{-1/2} = \frac{1}{\sqrt{\pi j}} + O(j^{-3/2})$$

in order to conclude that

$$[z^j]\frac{-\sqrt{(z-1)(z-4)(z^2+3z+4)}}{16(z-1)} = \frac{\sqrt{3}}{4\sqrt{2\pi j}} + O(j^{-1}).$$

We obtain

$$[z^j]\, H(z) = c_{t_j} = \frac{1}{2} + \frac{\sqrt{3}}{4\sqrt{2\pi j}} + O(j^{-1}).$$

This shows that $c_{t_j} > 1/2$ for sufficiently large j, which completes the proof of Theorem 3.3.

We note that by a closer look at the underlying proofs the error term $O(j^{-1})$ can be made effective so that is only necessary to check some initial values.

3.5 Proof of Theorem 3.4

3.5.1 The mean value of c_t

We want to compute the expected value of the sum of c_t over dyadic intervals,

$$S_\lambda = \frac{1}{2^\lambda} \sum_{2^\lambda \le t < 2^{\lambda+1}} c_t.$$

In order to find an explicit formula for S_λ, we introduce the more general expression

$$S_{k,\lambda} = \frac{1}{2^\lambda} \sum_{2^\lambda \le t < 2^{\lambda+1}} \delta(k,t).$$

We split into even and odd indices and observe that $\delta(k+1, 2^\lambda) = \delta(k+1, 2^{\lambda-1})$ to obtain

$$\begin{aligned} S_{k,\lambda} &= \frac{1}{2^\lambda} \sum_{2^{\lambda-1} \le t < 2^\lambda} \delta(k, 2t) + \frac{1}{2^\lambda} \sum_{2^{\lambda-1} \le t < 2^\lambda} \frac{1}{2} \left(\delta(k-1,t) + \delta(k+1, t+1) \right) \\ &= \frac{1}{4} \left(S_{k-1,\lambda-1} + 2S_{k,\lambda-1} + S_{k+1,\lambda-1} \right) \end{aligned}$$

for $\lambda \ge 1$. We perform λ steps of this recurrence to reduce $S_{k,\lambda}$ to a linear combination of the values $S_{l,0}$. By an induction involving binomial coefficients, we get for $0 \le \mu \le \lambda$

$$S_{k,\lambda} = \frac{1}{4^\mu} \sum_{s=-\mu}^{\mu} \binom{2\mu}{s+\mu} S_{k+s,\lambda-\mu}$$

and observing that

$$S_{k,0} = \delta(k,1) = \left\{ \begin{array}{cc} 2^{k-2}, & k \le 1 \\ 0 & \text{otherwise} \end{array} \right\},$$

we obtain

$$S_{k,\lambda} = \frac{1}{4^\lambda} \sum_{s=-\lambda}^{\lambda} \binom{2\lambda}{s+\lambda} \delta(k+s,1) = \frac{1}{4^\lambda} \sum_{s=0}^{2\lambda} \binom{2\lambda}{s} \delta(k+s-\lambda, 1) = \frac{1}{4^\lambda} \sum_{s=0}^{\lambda+1} \binom{2\lambda}{s} 2^{k+s-\lambda-2}$$

and therefore

$$\frac{1}{2^\lambda} \sum_{2^\lambda \le t < 2^{\lambda+1}} c_t = \sum_{k=0}^{\lambda} S_{k,\lambda} = \frac{1}{4^\lambda} \sum_{k=0}^{\lambda+1} \binom{2\lambda}{k} \left(1 - 2^{k-\lambda-2} \right).$$

We prove that this expression is always greater than $1/2$.

Proposition 3.6. *For all $\lambda \ge 0$ we have*

$$\frac{1}{2^\lambda} \sum_{2^\lambda \le t < 2^{\lambda+1}} c_t > \frac{1}{2}.$$

Moreover, for $\lambda \to \infty$ we have

$$\frac{1}{2^\lambda} \sum_{2^\lambda \le t < 2^{\lambda+1}} c_t = \frac{1}{2} + \frac{1}{2\sqrt{\pi\lambda}} + O\left(\lambda^{-3/2} \right).$$

Proof. We have

$$\sum_{s=0}^{\lambda+1} \binom{2\lambda}{s} \left(1 - 2^{s-\lambda-2} \right) \ge \sum_{s=0}^{\lambda-1} \binom{2\lambda}{s} + \frac{1}{2} \binom{2\lambda}{\lambda} + \frac{1}{2} \binom{2\lambda}{\lambda} + \binom{2\lambda}{\lambda+1} - 2^{-\lambda-2} \binom{2\lambda}{\lambda} \sum_{s=0}^{\lambda+1} 2^s$$

$$= \frac{1}{2}4^\lambda + \left(\frac{1}{2} - \frac{1}{\lambda+1} + O(2^{-\lambda})\right)\binom{2\lambda}{\lambda}$$

with a positive implied constant. For $\lambda \geq 2$ this expression is strictly greater than $\frac{1}{2}4^\lambda$. Moreover, the central binomial coefficient satisfies

$$\binom{2\lambda}{\lambda} = 4^\lambda\left(\frac{1}{\sqrt{\pi\lambda}} + O\left(\lambda^{-3/2}\right)\right).$$

□

As a corollary, we obtain that

$$\frac{1}{N}\sum_{t<N} c_t > \frac{1}{2}$$

for infinitely many N, which is a rather small result in view of the original problem. Nevertheless, Proposition 3.6 is going to be useful later on.

We note that the approach presented above does not carry over to arbitrary summation ranges easily.

3.5.2 A generating function for the second moment of c_t

Based on numerical experiments we expect that the standard deviation of c_t on dyadic intervals $[2^\lambda, 2^{\lambda+1}-1]$ is significantly smaller than $m_\lambda - 1/2$. If this holds true, an application of Chebychev's inequality yields $c_t > 1/2$ at least for a positive proportion of integers t (Actually we can prove more than that.) Our goal is therefore to find an explicit expression for the term

$$\sum_{2^\lambda \leq t < 2^{\lambda+1}} c_t^2.$$

We define

$$\begin{aligned}
a_{\lambda,k,\ell} &= 4^\lambda \sum_{2^\lambda \leq t < 2^{\lambda+1}} \delta(\lambda+1-k,t)\delta(\lambda+1-\ell,t)\\
b_{\lambda,k,\ell} &= 4^\lambda \sum_{2^\lambda \leq t < 2^{\lambda+1}} \delta(\lambda+1-k,t)\delta(\lambda+1-\ell,t+1)\\
c_{\lambda,k,\ell} &= 4^\lambda \sum_{2^\lambda \leq t < 2^{\lambda+1}} \delta(\lambda+1-k,t+1)\delta(\lambda+1-\ell,t).
\end{aligned}$$

We get for $\lambda \geq 1$

$$\begin{aligned}
&\sum_{2^\lambda \leq t < 2^{\lambda+1}} \delta(\lambda+1-k,t)\delta(\lambda+1-\ell,t)\\
&\qquad = \sum_{2^{\lambda-1} \leq t < 2^\lambda} \delta((\lambda-1)+1-(k-1),2t)\delta((\lambda-1)+1-(\ell-1),2t)\\
&\qquad + \sum_{2^{\lambda-1} \leq t < 2^\lambda} \delta(\lambda+1-k,2t+1)\delta(\lambda+1-\ell,2t+1)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{2^{\lambda-1}\le t<2^{\lambda}} \delta((\lambda-1)+1-(k-1),t)\delta((\lambda-1)+1-(\ell-1),t)\\
&+\frac{1}{4}\sum_{2^{\lambda-1}\le t<2^{\lambda}} (\delta(\lambda+1-k-1,t)+\delta(\lambda+1-k+1,t+1))\\
&\times(\delta(\lambda+1-\ell-1,t)+\delta(\lambda+1-\ell+1,t+1))\\
&= \sum_{2^{\lambda-1}\le t<2^{\lambda}} \delta((\lambda-1)+1-(k-1),t)\delta((\lambda-1)+1-(\ell-1),t)\\
&+\frac{1}{4}\sum_{2^{\lambda-1}\le t<2^{\lambda}} \delta(\lambda-1+1-k,t)\delta(\lambda-1+1-\ell,t)\\
&+\frac{1}{4}\sum_{2^{\lambda-1}\le t<2^{\lambda}} \delta(\lambda-1+1-(k-2),t+1)\delta(\lambda-1+1-(\ell-2),t+1)\\
&+\frac{1}{4}\sum_{2^{\lambda-1}\le t<2^{\lambda}} \delta(\lambda-1+1-k,t)\delta(\lambda-1+1-(\ell-2),t+1)\\
&+\frac{1}{4}\sum_{2^{\lambda-1}\le t<2^{\lambda}} \delta(\lambda-1+1-(k-2),t+1)\delta(\lambda-1+1-\ell,t)
\end{aligned}
$$

Therefore

$$a_{\lambda,k,\ell} = 4a_{\lambda-1,k-1,\ell-1} + a_{\lambda-1,k,\ell} + a_{\lambda-1,k-2,\ell-2} + b_{\lambda-1,k,\ell-2} + c_{\lambda-1,k-2,\ell}$$

Analogously, we have

$$
\begin{aligned}
&\sum_{2^{\lambda}\le t<2^{\lambda+1}} \delta(\lambda+1-k,t)\delta(\lambda+1-\ell,t+1)\\
&\qquad= \sum_{2^{\lambda-1}\le t<2^{\lambda}} \delta(\lambda+1-k,2t)\delta(\lambda+1-\ell,2t+1)\\
&\qquad+ \sum_{2^{\lambda-1}\le t<2^{\lambda}} \delta(\lambda+1-k,2t+1)\delta(\lambda+1-\ell,2t+2)\\
&\qquad= \frac{1}{2}\sum_{2^{\lambda-1}\le t<2^{\lambda}} \delta(\lambda+1-k,2t)\delta(\lambda+1-\ell-1,t)\\
&\qquad+ \frac{1}{2}\sum_{2^{\lambda-1}\le t<2^{\lambda}} \delta(\lambda+1-k,2t)\delta(\lambda+1-\ell+1,t+1)\\
&\qquad+ \frac{1}{2}\sum_{2^{\lambda-1}\le t<2^{\lambda}} \delta(\lambda+1-k-1,t)\delta(\lambda+1-\ell,t+1)\\
&\qquad+ \frac{1}{2}\sum_{2^{\lambda-1}\le t<2^{\lambda}} \delta(\lambda+1-k+1,t+1)\delta(\lambda+1-\ell,t+1),
\end{aligned}
$$

therefore

$$b_{\lambda,k,\ell} = 2a_{\lambda-1,k-1,\ell} + 2b_{\lambda-1,k-1,\ell-2} + 2b_{\lambda-1,k,\ell-1} + 2a_{\lambda-1,k-2,\ell-1}$$

for $\lambda \geq 1$ and $k, \ell \in \mathbb{Z}$. Finally, we calculate

$$\begin{aligned}
&\sum_{2^\lambda \leq t < 2^{\lambda+1}} \delta(\lambda+1-k, t+1)\delta(\lambda+1-\ell, t) \\
&\qquad = \sum_{2^{\lambda-1} \leq t < 2^\lambda} \delta(\lambda+1-k, 2t+1)\delta(\lambda+1-\ell, 2t) \\
&\qquad + \sum_{2^{\lambda-1} \leq t < 2^\lambda} \delta(\lambda+1-k, 2t+2)\delta(\lambda+1-\ell, 2t+1) \\
&\qquad = \frac{1}{2} \sum_{2^{\lambda-1} \leq t < 2^\lambda} \delta(\lambda+1-k-1, t)\delta(\lambda+1-\ell, 2t) \\
&\qquad + \frac{1}{2} \sum_{2^{\lambda-1} \leq t < 2^\lambda} \delta(\lambda+1-k+1, t+1)\delta(\lambda+1-\ell, 2t) \\
&\qquad + \frac{1}{2} \sum_{2^{\lambda-1} \leq t < 2^\lambda} \delta(\lambda+1-k, 2t+2)\delta(\lambda+1-\ell-1, t) \\
&\qquad + \frac{1}{2} \sum_{2^{\lambda-1} \leq t < 2^\lambda} \delta(\lambda+1-k, 2t+2)\delta(\lambda+1-\ell+1, t+1),
\end{aligned}$$

therefore

$$c_{\lambda,k,\ell} = 2a_{\lambda-1,k,\ell-1} + 2c_{\lambda-1,k-2,\ell-1} + 2c_{\lambda-1,k-1,\ell} + 2a_{\lambda-1,k-1,\ell-2}.$$

for $\lambda \geq 1$ and $k, \ell \in \mathbb{Z}$.

We define the trivariate generating functions

$$\begin{aligned}
F(x,y,z) &= \sum_{\lambda,k,\ell \geq 0} a_{\lambda,k,\ell} x^\lambda y^k z^\ell, \\
G(x,y,z) &= \sum_{\lambda,k,\ell \geq 0} b_{\lambda,k,\ell} x^\lambda y^k z^\ell, \\
H(x,y,z) &= \sum_{\lambda,k,\ell \geq 0} c_{\lambda,k,\ell} x^\lambda y^k z^\ell.
\end{aligned}$$

These are formal power series in three indeterminates. From the recurrence relations for a, b and c we get

$$\begin{aligned}
F(x,y,z) &= A + 4xyzF(x,y,z) + x(1+y^2z^2)F(x,y,z) + xz^2G(x,y,z) + xy^2H(x,y,z), \\
G(x,y,z) &= A + 2x(y+y^2z)F(x,y,z) + 2x(yz^2+z)G(x,y,z), \\
H(x,y,z) &= A + 2x(z+yz^2)F(x,y,z) + 2x(y^2z+y)H(x,y,z),
\end{aligned}$$

where

$$A = \sum_{k,\ell \geq 0} a_{0,k,\ell} x^0 y^k z^\ell.$$

We note that $a_{0,k,\ell} = b_{0,k,\ell} = c_{0,k,\ell}$. The equations for G and H can be written in the form

$$G(x,y,z) = \frac{A + 2xy(1+yz)F(x,y,z)}{1-2xz(1+yz)}.$$

and

$$H(x,y,z)=\frac{A+2xz(1+yz)F(x,y,z)}{1-2xy(1+yz)}$$

respectively. By inserting these two identities into the first equation we get

$$F(x,y,z)\left(1-4xyz-x(1+y^2z^2)-xz^2\frac{2xy(1+yz)}{1-2xz(1+yz)}-xy^2\frac{2xz(1+yz)}{1-2xy(1+yz)}\right)$$
$$=A\left(1+\frac{xz^2}{1-2xz(1+yz)}+\frac{xy^2}{1-2xy(1+yz)}\right).$$

We want to compute

$$\sum_{2^\lambda\le t<2^{\lambda+1}}c_t^2=\sum_{k,\ell\ge 0}\sum_{2^\lambda\le t<2^{\lambda-1}}\delta(k,t)\delta(\ell,t)$$
$$=\sum_{0\le k,\ell\le\lambda+1}\sum_{2^\lambda\le t<2^{\lambda-1}}\delta(\lambda+1-k,t)\delta(\lambda+1-\ell,t)=\frac{1}{4^\lambda}\sum_{k,\ell\le\lambda+1}a_{\lambda,k,\ell}.$$

Using this equality and

$$A=\sum_{k,\ell\ge 0}a_{0,k,\ell}x^0y^kz^\ell=\sum_{k,\ell\ge 0}\sum_{2^\lambda\le t<2^{\lambda+1}}\delta(1-k,t)\delta(1-\ell,t)x^0y^kz^\ell$$
$$=\sum_{k\ge 0}2^{-1-k}y^k\sum_{\ell\ge 0}2^{-1-\ell}z^\ell=\frac{1}{2-y}\frac{1}{2-z},$$

we get

$$\frac{1}{2^\lambda}\sum_{2^\lambda\le t<2^{\lambda+1}}c_t^2=\frac{1}{8^\lambda}\left[x^{\lambda+1}y^{\lambda+1}z^{\lambda+1}\right]\frac{x}{(1-y)(2-y)(1-z)(2-z)}$$
$$\times\frac{1+\frac{xz^2}{1-2xz(1+yz)}+\frac{xy^2}{1-2xy(1+yz)}}{1-4xyz-x(1+y^2z^2)-xyz\frac{2xz(1+yz)}{1-2xz(1+yz)}-xyz\frac{2xy(1+yz)}{1-2xy(1+yz)}}.\quad(3.13)$$

In the next section we will extract an asymptotic expansion for the diagonal sequence, which we are interested in.

3.5.3 Asymptotic expansion for the second moment of c_t

Set

$$F(x,y,z)=\frac{x}{(1-y)(2-y)(1-z)(2-z)}$$
$$\times\frac{1+\frac{xz^2}{1-2xz(1+yz)}+\frac{xy^2}{1-2xy(1+yz)}}{1-4xyz-x(1+y^2z^2)-xyz\frac{2xz(1+yz)}{1-2xz(1+yz)}-xyz\frac{2xy(1+yz)}{1-2xy(1+yz)}}.$$

The purpose of this section is to prove that the main diagonal of F satisfies the following asymptotic expansion.

Proposition 3.7. *The following asymptotic expansion holds as $n \to \infty$:*

$$[x^n y^n z^n]\, F(x,y,z) = 8^n \left(\frac{1}{32} + \frac{1}{16\sqrt{\pi}} \frac{1}{\sqrt{n}} + \frac{1}{32\pi} \frac{1}{n} + O(n^{-3/2}) \right).$$

Actually we will only derive the first two terms rigorously. The computation for the third term $8^n/(32\pi n)$ is a direct extension without any new ideas, but it would involve a very messy computation. For the sake of brevity we write $F(x,y,z)$ as

$$F(x,y,z) = \frac{x}{(1-y)(1-z)} \frac{G(x,y,z)}{H(x,y,z)},$$

where

$$G(x,y,z) = \frac{1 + \frac{xz^2}{1-2xz(1+yz)} + \frac{xy^2}{1-2xy(1+yz)}}{(2-y)(2-z)}$$

and

$$H(x,y,z) = 1 - 4xyz - x(1+y^2z^2) - xyz \frac{2xz(1+yz)}{1-2xz(1+yz)} - xyz \frac{2xy(1+yz)}{1-2xy(1+yz)}.$$

The idea of the proof is to extract first the n-th coefficient $[x^n]\, F(x,y,z)$ – which turns out to be easy because we just have a polar singularity in x – and then to apply the saddle point method in order to extract the coefficient $[x^n y^n z^n]\, F(x,y,z) = [y^n z^n]\, [x^n]\, F(x,y,z)$.

We start with the following property.

Lemma 3.8. *Let $f(y,z)$ be the unique solution of the equation*

$$H(f(y,z), y, z) = 0$$

with $f(1,1) = 1/8$. Then we have

$$[x^n]\, F(x,y,z) = \frac{1}{(1-y)(1-z)} \left(\frac{-G(f(y,z),y,z)}{H_x(f(y,z),y,z)}\, f(y,z)^{-n} + O(8^{(1-\epsilon)n}) \right) \tag{3.14}$$

uniformly for y, z in a complex neighborhood of 1, where $\epsilon > 0$.

Furthermore we have the local expansion

$$\begin{aligned} \log f(y,z) = &- \log 8 - (y-1) - (z-1) + \frac{1}{4}(y-1)^2 + \frac{1}{4}(z-1)^2 \\ &- \frac{1}{12}(y-1)^3 - \frac{1}{12}(z-1)^3 + O((y-1)^4 + (z-1)^4). \end{aligned}$$

Proof. Since $H(1/8,1,1) = 0$ and $H_x(1/8,1,1) = -12$ it follows from the implicit function theorem that there is a unique solution $x = f(y,z)$ of the equation $H(x,y,z) = 0$ with $f(1,1) = 1/8$. Furthermore it is an easy (but tedious) exercise of implicit differentiation to derive the local expansion of $f(y,z)$ and that of $\log f(y,z)$.

We now disregard the factor $(1-y)(1-z)$ in the denominator and suppose that y and z are contained in a sufficiently small (complex) neighborhood of 1. Then $x = f(y,z)$ is a polar singularity of the function $x \mapsto F(x,y,z)$ and it is easy to observe that there exists

$\epsilon > 0$ such that there is no other singularity for $|x| \le |f(y,z)| + \epsilon$. Consequently, using this property and by applying the local expansion, we obtain

$$(1-y)(1-z)F(x,y,z) = \frac{-G(f(y,z),y,z)}{H_x(f(y,z),y,z)} \frac{1}{1-x/f(y,z)} + O(1).$$

The asymptotic expansion (3.14) follows immediately by applying Cauchy's formula and the residue theorem. □

Note that the denominator $H(x,y,z)$ has the form $H(x,y,z) = 1 - P(x,y,z)$, where $P(x,y,z)$ considered as power series in x,y,z has only non-negative coefficients. Thus it follows that $H(x,y,z) \neq 0$ if $|x| \le 1/8$, $|y| \le 1$, $|z| \le 1$ but $y \neq 1$ or $z \neq 1$. This implies that there exist $\epsilon > 0$, $\delta_1 > 0$, and $\delta_2 > 0$ such that $|H(x,y,z)| \ge \epsilon$ for $|x| \le 1/8 + \delta_1$, $|y| \le 1 + \delta_1$, $|y| \le 1 + \delta_1$ but $|y-1| \ge \delta_2$ or $|z-1| \ge \delta_2$. In particular it follows that

$$[x^n]\, F(x,y,z) = O(8^{(1-\epsilon)n}) \tag{3.15}$$

if $|y| \le 1 + \delta_1$, $|y| \le 1 + \delta_1$ but $|y-1| \ge \delta_2$ or $|z-1| \ge \delta_2$.

The next lemma will be needed for computing the asymptotic expansion of the coefficients $[y^n z^n]$.

Lemma 3.9. *We have*

$$\int_{-\infty, \Re(s)>0}^{\infty} e^{-s^2/4} \frac{ds}{s} = -\pi i,$$

and

$$\int_{-\infty}^{\infty} e^{-s^2/4}\, ds = 2\sqrt{\pi}, \quad \int_{-\infty}^{\infty} e^{-s^2/4} s^2\, ds = 4\sqrt{\pi}.$$

Proof. Set

$$I = \int_{-\infty, \Re(s)>0}^{\infty} e^{-s^2/4} \frac{ds}{s}.$$

By substituting s by $-s$ it follows that

$$I = \int_{\infty, \Re(s)<0}^{-\infty} e^{-s^2/4} \frac{ds}{s}.$$

Hence, if we concatenate both integrals we encycle the origin clockwise so that the residue theorem implies

$$I + I = -2\pi i.$$

Consequently, $I = -\pi i$.

The remaining two integrals are standard Gaussian integrals. □

In order to determine the coefficient $[y^n z^n]$ we use Cauchy integration

$$[x^n y^n z^n] F(x,y,z) = \frac{1}{(2\pi i)^2} \iint_{\gamma \times \gamma} [x^n] F(x,y,z) \frac{dy}{y^{n+1}} \frac{dz}{z^{n+1}},$$

where γ is almost a cycle with radius 1. We just have to bypass $y = 1$ and $z = 1$, respectively.

By (3.14) and (3.15) we can replace γ by $\gamma' = \gamma \cap \{y \in \mathbb{C} : |y-1| \le \delta_2\}$ and obtain

$$[x^n y^n z^n]F(x,y,z) = \frac{1}{(2\pi i)^2} \iint\limits_{\gamma'\times\gamma'} \frac{1}{(1-y)(1-z)} \frac{-G(f(y,z),y,z)}{H_x(f(y,z),y,z)} f(y,z)^{-n} \frac{dy}{y^{n+1}} \frac{dz}{z^{n+1}} + O(8^{(1-\epsilon)n})$$

For $y, z \in \gamma'$ we set

$$y = 1 + i\frac{s}{\sqrt{n}} \quad \text{and} \quad z = 1 + i\frac{t}{\sqrt{n}}$$

and obtain

$$[x^n y^n z^n]F(x,y,z) = \frac{1}{(2\pi i)^2} \iint\limits_{|s|,|t|\le\delta_2\sqrt{n},\Re(s),\Re(t)>0} P_n(s,t) e^{-n\, g_n(s,t)} \frac{ds\, dt}{st} + O(8^{(1-\epsilon)n}),$$

where

$$P_n(s,t) = \left.\frac{-G(f(y,z),y,z)}{yzH_x(f(y,z),y,z)}\right|_{y=1+is/\sqrt{n},\, z=1+it/\sqrt{n}}$$

and

$$g_n(s,t) = \left.(f(y,z) + \log y + \log z)\right|_{y=1+is/\sqrt{n},\, z=1+it/\sqrt{n}}.$$

Since

$$\frac{-G(f(y,z),y,z)}{yzH_x(f(y,z),y,z)} = \frac{1}{8} - \frac{1}{8}(y-1) - \frac{1}{8}(z-1) + O((y-1)^2 + (z-1)^2)$$

it follows that

$$P_n(s,t) = \frac{1}{8}\left(1 - \frac{is}{\sqrt{n}} - \frac{it}{\sqrt{n}} + O\left(\frac{s^2+t^2}{n}\right)\right).$$

Lemma 3.8 implies

$$\log f(y,z) + \log y + \log z = -\log 8 - \frac{1}{4}(y-1)^2 - \frac{1}{4}(z-1)^2 + \frac{1}{4}(y-1)^3 + \frac{1}{4}(z-1)^3 + O((y-1)^4 + (z-1)^4)$$

so that

$$\begin{aligned} -n\, g_n(s,t) &= n\left(\log 8 + \frac{1}{4}(y-1)^2 + \frac{1}{4}(z-1)^2 - \frac{1}{4}(y-1)^3 - \frac{1}{4}(z-1)^3\right) \\ &\quad + O(n(y-1)^4 + n(z-1)^4) \\ &= \log 8^n - \frac{s^2}{4} - \frac{t^2}{4} + i\frac{s^3}{4\sqrt{n}} + i\frac{t^3}{4\sqrt{n}} + O\left(\frac{s^4+t^4}{n}\right) \end{aligned}$$

and

$$e^{-n\, g_n(s,t)} = 8^n e^{-\frac{s^2}{4}-\frac{t^2}{4}}\left(1 + i\frac{s^3}{4\sqrt{n}} + i\frac{t^3}{4\sqrt{n}} + O\left(\frac{s^4+s^6+t^4+t^6}{n}\right)\right).$$

This leads to

$$\begin{aligned}
&\frac{1}{(2\pi i)^2} \iint\limits_{|s|,|t|\le\delta_2\sqrt{n},\Re(s),\Re(t)>0} P_n(s,t)e^{-n\,g_n(s,t)}\,\frac{ds\,dt}{st}\\
&= \frac{8^{n-1}}{(2\pi i)^2} \iint\limits_{|s|,|t|\le\delta_2\sqrt{n},\Re(s),\Re(t)>0} e^{-\frac{s^2}{4}-\frac{t^2}{4}}\left(1+i\frac{s^3/4-s}{\sqrt{n}}+i\frac{t^3/-t}{\sqrt{n}}\right)\frac{ds\,dt}{st}\\
&\quad + O\left(\frac{8^n}{n}\right)\\
&= \frac{8^{n-1}}{(2\pi i)^2} \iint\limits_{-\infty<s,t<\infty,\Re(s),\Re(t)>0} e^{-\frac{s^2}{4}-\frac{t^2}{4}}\left(1+i\frac{s^3/4-2s}{\sqrt{n}}+i\frac{t^3/4-t}{\sqrt{n}}\right)\frac{ds\,dt}{st}\\
&\quad + O\left(\frac{8^n}{n}\right)
\end{aligned}$$

Finally by applying Lemma 3.9 we obtain

$$\begin{aligned}
&\frac{1}{(2\pi i)^2} \iint\limits_{-\infty<s,t<\infty,\Re(s),\Re(t)>0} e^{-\frac{s^2}{4}-\frac{t^2}{4}}\left(1+i\frac{s^3/4-s}{\sqrt{n}}+i\frac{t^3/4-t}{\sqrt{n}}\right)\frac{ds\,dt}{st}\\
&= \frac{1}{(2\pi i)^2}\left((-i\pi)^2+2(i4\sqrt{\pi}(-i\pi)/4-2\sqrt{\pi}(-i\pi))/\sqrt{n}\right)\\
&= \frac{1}{4}+\frac{1}{2\sqrt{\pi n}}
\end{aligned}$$

Summing up this implies

$$\begin{aligned}
[x^n y^n z^n]F(x,y,z) &= 8^{n-1}\left(\frac{1}{4}+\frac{1}{2\sqrt{\pi n}}+O\left(\frac{1}{n}\right)\right)\\
&= 8^n\left(\frac{1}{32}+\frac{1}{16\sqrt{\pi n}}+O\left(\frac{1}{n}\right)\right)
\end{aligned}$$

and proves the first two terms in the asymptotic expansion of Proposition 3.7.

By inspecting the above proof it is clear that there is an asymptotic series expansion of the form

$$[x^n y^n z^n]F(x,y,z) \sim 8^n\left(c_0+\frac{c_1}{\sqrt{n}}+\frac{c_2}{n}+\frac{c_3}{n^{3/2}}+\cdots\right)$$

with real constants c_j. It is just a computational question to determine them. In particular it follows that $c_2 = 1/(32\pi)$ which completes the proof of Proposition 3.7.

3.5.4 Completing the proof of Theorem 3.4

From Proposition 3.7 and (3.13) we obtain

$$\frac{1}{2^\lambda}\sum_{2^\lambda\le t<2^{\lambda+1}} c_t^2 = \frac{1}{8^\lambda}\left[x^{\lambda+1}y^{\lambda+1}z^{\lambda+1}\right]F(x,y,z)$$

$$= \frac{1}{8^\lambda} 8^{\lambda+1} \left(\frac{1}{32} + \frac{1}{16\sqrt{\pi}} \frac{1}{\sqrt{\lambda+1}} + \frac{1}{32\pi} \frac{1}{\lambda+1} + O((\lambda+1)^{-3/2}) \right).$$

By the mean value theorem we have $(\lambda+1)^a = \lambda^a + O(\lambda^{a-1})$ for $a < 0$ and $\lambda > 0$, therefore

$$\frac{1}{2^\lambda} \sum_{2^\lambda \le t < 2^{\lambda+1}} c_t^2 = \frac{1}{4} + \frac{1}{2\sqrt{\pi}} \lambda^{-1/2} + \frac{1}{4\pi} \lambda^{-1} + O(\lambda^{-3/2}).$$

On the other hand Proposition 3.6 implies

$$\left(\frac{1}{2^\lambda} \sum_{2^\lambda \le t < 2^{\lambda+1}} c_t \right)^2 = \frac{1}{4} + \frac{1}{2\sqrt{\pi}} \lambda^{-1/2} + \frac{1}{4\pi} \lambda^{-1} + O(\lambda^{-3/2}).$$

From these formulas it follows that the standard deviation σ_λ of the discrete random variable c_t (where $t \in [2^\lambda, 2^{\lambda+1})$) satisfies

$$\sigma_\lambda = O(\lambda^{-3/4}).$$

Since the sequence of standard deviations converges to zero faster than the sequence of distances of the expected values from $1/2$, the theorem of Chebyshev can be applied to yield the density 1-result. More precisely, let X_λ be the sequence c_t on the interval $[2^\lambda, 2^{\lambda+1})$, viewed as a random variable. Since $\mathbb{E}X_\lambda > 1/2$, we obtain

$$\begin{aligned}
\mathbb{P}(X_\lambda > 1/2) &\ge \mathbb{P}\left(|X_\lambda - \mathbb{E}X_\lambda| < \frac{\mathbb{E}X_\lambda - 1/2}{\sigma_\lambda} \sigma_\lambda \right) \\
&\ge 1 - \left(\frac{\mathbb{E}X_\lambda - 1/2}{\sigma_\lambda} \right)^{-2} \\
&\gg 1 - \left(\frac{\lambda^{-1/2}}{\lambda^{-3/4}} \right)^{-2} \\
&\gg 1 - \lambda^{-1/2}
\end{aligned}$$

as $\lambda \to \infty$. For $\mu \ge 0$ we obtain therefore

$$\begin{aligned}
2^\mu - |\{t < 2^\mu : c_t > 1/2\}| &= 2^\mu - \sum_{\lambda < \mu} \left| \{t \in \{2^\lambda, \dots, 2^{\lambda+1} - 1\} : c_t > 1/2\} \right| \\
&\ll 2^\mu - \sum_{\lambda < \mu} 2^\lambda \left(1 - \lambda^{-1/2}\right) \\
&\ll \sum_{\lambda < \mu} 2^\lambda \lambda^{-1/2} \\
&\ll \sum_{\lambda < \mu'} 2^\lambda + \sum_{\mu' \le \lambda < \mu} 2^\lambda \mu'^{-1/2} \\
&\ll 2^{\mu'} + 2^\mu \mu'^{-1/2} \\
&\ll 2^\mu \mu^{-1/2},
\end{aligned}$$

where $\mu' = \mu/2$ (for example). Let $T > 0$ be arbitrary and $2^\mu \le T < 2^{\mu+1}$. Then

$$\begin{aligned} T - |\{t < T : c_t > 1/2\}| &\le 2^\mu - |\{t < 2^\mu : c_t > 1/2\}| + |\{t \in \{2^\mu, \dots, T-1\} : c_t \le 1/2\}| \\ &\ll 2^\mu \mu^{-1/2} + \left|\{t \in \{2^\mu, \dots, 2^{\mu+1}-1\} : c_t \le 1/2\}\right| \\ &\ll 2^\mu \mu^{-1/2} + 2^\mu - \left|\{t \in \{2^\mu, \dots, 2^{\mu+1}-1\} : c_t > 1/2\}\right| \\ &\ll 2^\mu \mu^{-1/2} \\ &\ll \frac{T}{\sqrt{\log T}}. \end{aligned}$$

This finishes the proof of Theorem 3.4.

3.6 Remarks on the generating function approach

We finish this chapter with a remark on a generating function approach to the double family δ, defined by the recurrence relation (3.4). We want to take the values $\delta(k,t)$ as coefficients of a bivariate generating function. Since we want to work with power series and avoid Laurent series, we have to take some care, since $\delta(k,t)$ can be nonzero for arbitrarily large k. However, if t is restricted to an interval $\{2^\lambda, \dots, 2^{\lambda+1}-1\}$, we have $\delta(k,t) = 0$ for $k > \lambda+1$. We define therefore for $\lambda \ge 0$ the bivariate generating function

$$F_\lambda(x,y) = \sum_{\substack{k \ge 0 \\ 2^\lambda \le t < 2^{\lambda+1}}} \delta(\lambda+1-k,t) x^{t-1} y^k, \tag{3.16}$$

which captures all "interesting" (that is, nonzero) values of $\delta(k,t)$. In particular, c_t can be recovered from this function, more precisely

$$c_t = \sum_{j \le \lambda+1} \left[x^{t-1} y^k\right] F_\lambda(x,y).$$

Using the recurrence relation, we obtain for $\lambda \ge 1$ after some straightforward calculation

$$F_\lambda(x,y) = \frac{1}{2}(x+y)^2 F_{\lambda-1}(x^2,y) + \frac{1}{2} y^{\lambda+1} A\left(x^{2^{\lambda+1}-2} - x^{2^\lambda-2}\right),$$

where

$$A = \sum_{k \ge 0} \delta(1-k,1) y^k.$$

Iterating this identity λ times, we obtain

$$F_\lambda(x,y) = \frac{A}{2}\left(x^{2^\lambda}-1\right) \sum_{0 \le j < \lambda} y^{\lambda+1-j} x^{2^\lambda - 2^{j+1}} \frac{1}{2^j} \prod_{0 \le i < j} \left(x^{2^i}+y\right)^2 + \frac{A}{2^\lambda} \prod_{0 \le i < \lambda} \left(x^{2^i}+y\right)^2.$$

We can therefore prove something on the values c_t as soon as we understand the coefficients of the product

$$\tilde{P}_j(x,y) = \prod_{i<j} \left(x^{2^i}+y\right)^2, \tag{3.17}$$

which is the same as understanding the product

$$P_j(x,y) = \prod_{i<j} \left(1 + x^{2^i} y\right)^2. \tag{3.18}$$

(By induction on j we can show that the corresponding coefficients $a_{m,n} = [x^n y^m] P_j(x,y)$ and $\tilde{a}_{m,n} = [x^n y^m] \tilde{P}_j(x,y)$ satisfy $a_{m,n} = \tilde{a}_{2j-m,n}$ for $0 \le m \le 2j$ and $n \ge 0$.) We list the arrays of coefficients of the first few products. Note that each array is constructed by

	$[x^0]$
$[y^0]$	1

Table 3.1: Coefficients for $j = 0$

	$[x^0]$	$[x^1]$	$[x^2]$
$[y^0]$	1		
$[y^1]$		2	
$[y^2]$			1

Table 3.2: Coefficients for $j = 1$

	$[x^0]$	$[x^1]$	$[x^2]$	$[x^3]$	$[x^4]$	$[x^5]$	$[x^6]$
$[y^0]$	1						
$[y^1]$		2	2				
$[y^2]$			1	4	1		
$[y^3]$					2	2	
$[y^4]$							1

Table 3.3: Coefficients for $j = 2$

	$[x^0]$	$[x^1]$	$[x^2]$	$[x^3]$	$[x^4]$	$[x^5]$	$[x^6]$	$[x^7]$	$[x^8]$	$[x^9]$	$[x^{10}]$	$[x^{11}]$	$[x^{12}]$	$[x^{13}]$	$[x^{14}]$
$[y^0]$	1														
$[y^1]$		2	2		2										
$[y^2]$			1	4	1	4	4		1						
$[y^3]$					2	2	2	8	2	2	2				
$[y^4]$							1		4	4	1	4	1		
$[y^5]$											2		2	2	
$[y^6]$															1

Table 3.4: Coefficients for $j = 3$

summing shifted and copies of the previous one, which corresponds to multiplication by the factor $1 + 2x^{2^i} y + x^{2^{i+1}} y^2$. The structure of the coefficients of the first few $P_j(x,y)$ is not

$[y^0]$	1																														
$[y^1]$		2	2		2				2																						
$[y^2]$			1	4	1	4	4		1	4	4		4				1														
$[y^3]$					2	2	2	8	2	2	4	8	2	8	8		2	2	2		2										
$[y^4]$							1		4	4	1	4	5	4	4	16	4	4	5	4	1	4	4		1						
$[y^5]$											2		2	2	2		8	8	2	8	4	2	2	8	2	2	2				
$[y^6]$															1				4		4	4	1		4	4	1	4	1		
$[y^7]$																							2				2		2	2	
$[y^8]$																															1

Table 3.5: Coefficients for $j = 4$

hard to understand, since the "building blocks" are placed disjointly to each other. However, beginning at $j = 4$ the patterns begin to overlap, which complicates matters.

It is easy to see that the sequence of products P_j converges to a bivariate power series

$$P(x,y) = \prod_{i \geq 0} \left(1 + x^{2^i} y\right)^2 \in \mathbb{Z}[x,y].$$

The double family of coefficients of this power series contains "the first half" of the double family of coefficients of of each finite product $P_j(x,y)$ as an initial segment. It is therefore sufficient to study this bivariate power series in order to understand the values c_t. However, while this looks like a feasible task, this product representation is not easy to handle.

The double sequence $\left([x^m y^n]P(x,y)\right)_{m,n\geq 0}$ is constructed in a similar way as the Stern-Brocot sequence. We recall that the elements of this sequence appear as the coefficients of the infinite product

$$\prod_{i \geq 0} \left(1 + x^{2^i} + x^{2^{i+1}}\right)$$

(see (2.29), compare also to formulas (2.28) and (2.30), in which products of a similar form feature). We start with the one-element sequence (1). In step i, the current sequence of coefficients is shifted by 0, 2^i and 2^{i+1} places respectively and the results are added componentwise. We obtain the sequences $(1,1,1)$, $(1,1,2,1,2,1,1)$, $(1,1,2,1,3,2,3,1,3,2,3,1,2,1,1)$, $(1,1,2,1,3,2,3,1,4,3,5,2,5,3,4,1,4,3,5,2,5,3,4,1,3,2,3,1,2,1,1)$ etc., converging componentwise to the Stern-Brocot sequence.

As it turns out, even this (univariate) sequence a is not easy to grasp. For example, it satisfies the unexpected property that $n \mapsto a_n/a_{n+1}$ is a *bijection* from the set of nonnegative integers to the set of positive rationals. We refer again to Northshield [48] for a proof of this and other properties of the Stern-Brocot sequence. In view of the apparent difficulties connected to this sequence, we have to expect that our product $P(x,y)$, which is defined in a similar way, does not reveal all of its secrets voluntarily and therefore cannot (yet) be used for deriving an answer to Question 1.

Chapter 4

Piatetski-Shapiro sequences via Beatty sequences

Integer sequences of the form $\lfloor n^c \rfloor$, where $1 < c < 2$, can be locally approximated by sequences of the form $\lfloor n\alpha + \beta \rfloor$ in a very good way. Following this approach, we are led to an estimate of the difference

$$\sum_{n \le x} \varphi\left(\lfloor n^c \rfloor\right) - \frac{1}{c} \sum_{n \le x^c} \varphi(n) n^{\frac{1}{c}-1},$$

which measures the deviation of the mean value of φ on the subsequence $\lfloor n^c \rfloor$ from the expected value, by an expression involving exponential sums. As an application we prove that for $1 < c \le 1.42$ the subsequence of the Thue-Morse sequence indexed by $\lfloor n^c \rfloor$ attains both of its values with asymptotic density $1/2$.

4.1 Introduction

Piatetski-Shapiro sequences are sequences of the form $(\lfloor n^c \rfloor)_{n \ge 1}$, where $c > 1$ is not an integer. They are named after I. Piatetski-Shapiro, who proved the following Prime Number Theorem (see [51]): If $1 < c < \frac{12}{11}$, then

$$|\{n \le x : \lfloor n^c \rfloor \text{ is prime}\}| \sim \frac{x}{c \log x}. \tag{4.1}$$

The range for c has been extended several times, the currently best known upper bound being $c < \frac{2817}{2426}$ obtained by Rivat and Sargos [53]. It is expected that the asymptotic formula (4.1) holds for all $c \in (1, 2)$, an expectation that is backed up by the fact that it is true for almost all $c \in [1, 2]$ with respect to the Lebesgue measure (see [36]).

For a collection of various arithmetic results on Piatetski-Shapiro sequences see the article [1] by Baker et al. For example in that article it is proved in detail that for $1 < c < \frac{149}{87}$ the number of squarefree integers of the form $\lfloor n^c \rfloor$ behaves as expected: for c in this range we have

$$|\{n \le x : \lfloor n^c \rfloor \text{ is squarefree}\}| = \frac{6}{\pi^2} x + O\left(x^{1-\varepsilon}\right).$$

According to that paper, this result was sketched by Cao and Zhai [8] before.

A more basic question is to ask for the distribution of $\lfloor n^c \rfloor$ in residue classes. In this case it is known that for all noninteger $c > 1$, all positive integers m and all $a \in \mathbb{Z}$ we have

$$|\{n \leq x : \lfloor n^c \rfloor \equiv a \bmod m\}| = \frac{x}{m} + O\left(x^{1-\varepsilon}\right)$$

for some $\varepsilon = \varepsilon(c)$ that can be given explicitly, see Deshouillers [17] and Morgenbesser [46].

Another line of research was initiated by Mauduit and Rivat [39] which concerns the behaviour of q-multiplicative functions on Piatetski-Shapiro sequences. For an integer $q \geq 2$, a function $\varphi : \mathbb{N} \to \mathbb{C}$ is called *q-multiplicative* if for all $a \geq 0$, $k \geq 0$ and for $0 \leq b < q^k$ we have $\varphi\left(q^k a + b\right) = \varphi\left(q^k a\right)\varphi(b)$. The function $\mathrm{e}\left(\alpha s_q(n)\right)$, where s_q denotes the sum-of-digits function in base q, and the trigonometric monomial $\mathrm{e}\left(\alpha n\right)$ are examples of q-multiplicative functions. Gelfond [29] solved the problem of describing the distribution of the values $s_q(n)$ in residue classes, where n itself is restricted to a residue class, and posed the analogous problem of describing the distribution of $s_q(P(n))$ in residue classes, where P is a polynomial of degree greater than one such that $P(\mathbb{N}) \subseteq \mathbb{N}$. The study of q-multiplicative functions on Piatetski-Shapiro sequences can be seen as a step towards the resolution of this question, in the same way that the Piatetski-Shapiro Prime Number Theorem is an approach to unsolved problems such as proving that there are infinitely many prime numbers of the form $n^2 + 1$. In [40] Mauduit and Rivat proved the following theorem.

Theorem A (Mauduit and Rivat)**.** *Let $c \in (1, 7/5)$ and $\gamma = 1/c$. For all $\delta \in (0, (7 - 5c)/9)$ there exists a constant $C = C(\gamma, \delta)$ such that for all q-multiplicative functions χ and all $x \geq 1$ we have*

$$\left| \sum_{1 \leq n \leq x} \chi\left(\lfloor n^c \rfloor\right) - \sum_{1 \leq m \leq x^c} \gamma m^{\gamma - 1} \chi(m) \right| \leq C(\gamma, \delta) x^{1-\delta}. \tag{4.2}$$

Morgenbesser [46] gave a nontrivial bound for the sum $\sum \mathrm{e}\left(\alpha s_q\left(\lfloor n^c \rfloor\right)\right)$ for all noninteger $c > 1$, provided only that q is large enough (depending on c). Deshouillers, Drmota and Morgenbesser [18] investigated subsequences of automatic sequences of the form $\lfloor n^c \rfloor$ for $c < 7/5$ by generalizing the method from [40]. Mauduit and Rivat [41] gave a complete description of the distribution of the sum of digits of squares in residue classes, thus solving the Gelfond problem concerning polynomials for the case that $P(X) = X^2$. The problem of proving (4.2) for the case that $c \geq 7/5$ is not an integer, $\chi(n) = \mathrm{e}\left(\alpha s_q(n)\right)$ and q is small could not be solved, however.

In the present article we follow a new approach to problems on Piatetski-Shapiro sequences. This approach is based on the idea of approximating the function x^c by a family of tangents $x\alpha + \beta$, each restricted to a small interval. Let $\delta \in (0, 1 - c/2)$ and $\varepsilon > 0$ be given. Then by linear approximation we can choose for $x_0 \geq 1$ some α and β in such a way that $|x^c - x\alpha - \beta| < \varepsilon$ if $|x - x_0| < Cx^\delta$, where C does not depend on x_0. It seems therefore likely that $\lfloor n^c \rfloor = \lfloor n\alpha + \beta \rfloor$ for most integers n in such an interval. These observations are made precise by the lemmas in Section 4.4.1.

Algebraic properties of the function $x \mapsto x^c$ are not needed for such an approximation. Correspondingly our method can be adapted to treat functions from a larger class, defined by certain conditions on the derivatives. Functions like $x^c \log^\eta x$ or $x^c \exp\left(\log^\varepsilon x\right)$, where $1 < c < 2$, $\eta \in \mathbb{R}$ and $0 \leq \varepsilon < 1$, are contained in this class as well as linear combinations with positive coefficients of its elements.

A sequence of integers of the form $(\lfloor n\alpha+\beta\rfloor)_{n\geq 1}$, where $\alpha > 0$, is called a *(non-homogeneous) Beatty sequence*. They are named after S. Beatty, who posed a problem (concerning the homogeneous case) in the American Mathematical Monthly in 1926 (see [4]), which essentially states that for irrational $\alpha_1, \alpha_2 > 1$ such that $\frac{1}{\alpha_1}+\frac{1}{\alpha_2}=1$ the sequences $(\lfloor n\alpha_1\rfloor)_{n\geq 1}$ and $(\lfloor n\alpha_2\rfloor)_{n\geq 1}$ form a partition of the set of positive integers. This fact was already found in 1894 by Rayleigh [52, pp.122–123] and correspondingly it is called Rayleigh's Theorem or Beatty's Theorem. We refer to [2] for some references to the newer literature concerning Beatty sequences.

We consider a bounded arithmetic function φ and a differentiable function $f:\mathbb{R}^+\to\mathbb{R}^+$ satisfying $f'>0$ and other conditions on its derivatives and ask whether it is true that

$$\sum_{A<n\leq 2A}\varphi\left(\lfloor f(n)\rfloor\right)-\sum_{f(A)<m\leq f(2A)}\varphi(m)\left(f^{-1}\right)'(m)=o(A) \tag{4.3}$$

as $A\to\infty$. The two terms on the left hand side resemble the terms involved in the change of variables in an integral. Heuristically, we expect therefore that "well behaved" functions φ yield a small error term on the right hand side. This expectation is in general very difficult to verify, which is obvious from the observation that, for instance, (4.1) can be reduced to a statement of the form (4.3).

The main result of this paper, based on the method of approximating $\lfloor n^c\rfloor$ by Beatty sequences and the approximation of the periodic Bernoulli polynomial $\psi(x)=x-\lfloor x\rfloor-\frac{1}{2}$ by trigonometric polynomials, is a sufficient condition for the statement (4.3) to hold. More precisely we give an upper bound on the error term that involves the exponential sum $\sum\varphi(m)e(m\vartheta)$ over short intervals.

We give several application of this theorem. The first application is an improvement of the bound $7/5=1.4$ in Theorem A to the value 1.42 in the case that χ is the Thue-Morse sequence, which expresses the parity of the number of ones in the binary representation of a natural number. In order to prove this result, we use an estimate of the L^1-norm of the corresponding exponential sum (as a function in ϑ) given by Fouvry and Mauduit [27].

Another application concerns the joint distribution of sum-of-digits functions on Piatetski-Shapiro sequences. It is another problem posed in the paper [29] by Gelfond to prove that if $q_1,q_2\geq 2$, $m_1,m_2\geq 1$ and l_1,l_2 are integers such that $(q_1,q_2)=1$, $(m_1,q_1-1)=1$ and $(m_2,q_2-1)=1$, there exists $\varepsilon>0$ such that

$$\begin{aligned}&\left|\{n\leq x: s_{q_1}(n)\equiv l_1 \bmod m_1 \text{ and } s_{q_2}(n)\equiv l_2 \bmod m_2\}\right|\\ &\qquad\qquad=\frac{x}{m_1m_2}+O\left(x^{1-\varepsilon}\right).\end{aligned} \tag{4.4}$$

This statement was proved by Kim [33], but a weaker form of this result, specifically with a non-explicit error term, was provided by Bésineau long before (see [7]). To the author's knowledge the problem of proving a result such as (4.4) for subsequences $\lfloor n^c\rfloor$ of the integers has not been dealt with in the literature before. We obtain such a result for all c in the interval $(1,18/17)$. In the proof we make (besides the main theorem) use of discrete Fourier coefficients related to the sum-of-digits function. These Fourier coefficients have proven to be an excellent tool for treating problems related to the sum of digits (see [41, 42]) and can also be used in this context. We also note that their use leads to an alternative method of proving (4.4).

As the third application we prove a result on the distribution in residue classes of the Zeckendorf sum-of-digits function s_Z evaluated on Piatetski-Shapiro sequences. By the well-known theorem of Zeckendorf [57] every positive integer n can be represented uniquely as a sum of non-consecutive Fibonacci numbers. The number of summands in this representation is called the *Zeckendorf sum-of-digits* of n, which we denote by $s_Z(n)$. We prove that for integers $m \geq 1$ and a and for all $c \in (1, 4/3)$ there exists $\varepsilon > 0$ such that

$$|\{n \leq x : s_Z(\lfloor n^c \rfloor) \equiv a \bmod m\}| = \frac{x}{m} + O\left(x^{1-\varepsilon}\right).$$

In this article, we denote the set of positive real numbers by $\mathbb{R}^+$ and the set of nonnegative integers by $\mathbb{N}$. For $x \in \mathbb{R}$ we write $\mathrm{e}(x) = \mathrm{e}^{2\pi i x}$, $\|x\| = \min_{n\in\mathbb{Z}} |n - x|$ and $\{x\} = x - \lfloor x \rfloor$. Conditions like $i < n$ under a summation or product sign are to be read as $0 \leq i < n$.

4.2 Main results

The main result is an estimate of the error term in (4.3) for a special class of functions f.

Theorem 4.1. *Assume that f is a two times continuously differentiable real valued function on $\mathbb{R}^+$ such that $f, f', f'' > 0$ and that there exist $c_1 \geq 1/2$ and $c_2 > 0$ such that for $0 < x \leq y \leq 2x$ we have $c_1 f''(x) \leq f''(y) \leq c_2 f''(x)$. Let $A_0 \geq 2$ be such that $f'(A_0) \geq 1$. There exists a constant $C = C(f)$ such that for all complex valued arithmetic functions φ bounded by 1, for all integers $A \geq A_0$ and for all $z > 0$ we have*

$$\frac{1}{A}\left|\sum_{A<n\leq 2A} \varphi\left(\lfloor f(n)\rfloor\right) - \sum_{f(A)<m\leq f(2A)} \varphi(m)\left(f^{-1}\right)'(m)\right| \leq C\left(\frac{f''(A)}{f'(A)^2}z^2 + f'(A)(\log A)^3 J(A,z)\right), \quad (4.5)$$

where

$$J(A,z) = \int_0^1 \sup_{f(A)<x\leq f(2A)} \frac{1}{z}\left|\sum_{x<m\leq x+z} \varphi(m)\,\mathrm{e}\left(m\vartheta\right)\right| \mathrm{d}\vartheta. \quad (4.6)$$

Theorem 4.1 is a consequence of the following result, which provides a way to prove a discrete substitution rule by solving a problem about the behaviour of φ on Beatty sequences.

Proposition 4.2. *Assume that f is a two times continuously differentiable real valued function on $\mathbb{R}^+$ such that $f, f', f'' > 0$, and that there exist $c_1 \geq 1/2$ and $c_2 > 0$ such that for $0 < x \leq y \leq 2x$ we have $c_1 f''(x) \leq f''(y) \leq c_2 f''(x)$. There exists $C = C(f)$ such that for all complex valued arithmetic functions φ bounded by 1, for all $A \geq 2$ and $K > 0$ we have*

$$\frac{1}{A}\left|\sum_{A<n\leq 2A} \varphi\left(\lfloor f(n)\rfloor\right) - \sum_{f(A)<m\leq f(2A)} \varphi(m)\left(f^{-1}\right)'(m)\right| \leq C\left(f''(A)K^2 + \frac{(\log A)^2}{K} + I(A,K)\right), \quad (4.7)$$

where $I(A,K)$ is defined by

$$I(A,K)=\frac{1}{f'(2A)-f'(A)}\times\int_{f'(A)}^{f'(2A)}\sup_{f(A)<\beta\le f(2A)}\frac{1}{K}\left|\sum_{0<n\le K}\varphi\left(\lfloor n\alpha+\beta\rfloor\right)-\frac{1}{\alpha}\sum_{\beta<m\le\beta+K\alpha}\varphi(m)\right|\mathrm{d}\alpha. \quad (4.8)$$

4.3 Applications

In the proofs of our applications, concerning sum-of-digits functions, we make use of bounds for the exponential sum $\sum_{x<m\le x+z}\varphi(m)\,\mathrm{e}\left(m\vartheta\right)$ that are independent of the value of x. Moreover, for simplicity we concentrate on the case that $f(x)=x^c$, although it would be possible to derive analogous results for a larger class of functions, as we noted in the introduction. We state a corollary of Theorem 4.1 that is adjusted to this situation.

Corollary 4.3. *Let φ be a complex valued arithmetic function bounded by 1. If $a\in(0,1]$ and C are such that*

$$\int_0^1\sup_{x\ge0}\left|\sum_{x<m\le x+z}\varphi(m)\,\mathrm{e}\left(m\vartheta\right)\right|\mathrm{d}\vartheta\le Cz^a \quad (4.9)$$

for $z\ge1$, then for all $c\in(1,2)$ and all $\eta\in\left(0,\frac{2-(a+1)c}{3-a}\right)$ there is a $C_1=C_1(a,c,C,\eta)$ such that

$$\frac{1}{N}\left|\sum_{1\le n\le N}\varphi\left(\lfloor n^c\rfloor\right)-\frac{1}{c}\sum_{1\le m\le N^c}\varphi(m)m^{\frac{1}{c}-1}\right|\le C_1N^{-\eta} \quad (4.10)$$

for $N\ge1$.

Proof. For $A>0$ we write

$$F(A)=\left|\sum_{A<n\le2A}\varphi\left(\lfloor n^c\rfloor\right)-\frac{1}{c}\sum_{A^c<m\le(2A)^c}\varphi(m)m^{\frac{1}{c}-1}\right|. \quad (4.11)$$

Let $1<c<2$ and set $z=A^{\frac{2c-1}{3-a}}$ for $A\ge2$. From hypothesis (4.9) and Theorem 4.1 it follows by a short calculation that for all integers $A\ge2$ and all $\varepsilon>0$ we have

$$F(A)\ll A^{1-\rho+\varepsilon} \quad (4.12)$$

with the choice $\rho=\frac{2-c(a+1)}{3-a}$. The implied constant in (4.12) may depend on a,c,C and ε. Altering the summation limits in (4.11) to $\lfloor A\rfloor<n\le\lfloor2A\rfloor$ and $\lfloor A\rfloor^c<m\le\lfloor2A\rfloor^c$ respectively introduces an error term of $O(1)$, which is neglegible. Therefore (4.12) holds for all real $A\ge2$ and $\varepsilon>0$. We have $F(A)=0$ for $A<\frac{1}{2}$, and it is clear that $F(A)$ is bounded for $0<A\le2$. From these observations and (4.12) it follows that $F(A)\ll A^{1-\rho+\varepsilon}$ for all $A>0$. Since $\rho-\varepsilon<1$ we get

$$\left|\sum_{1\le n\le N}\varphi\left(\lfloor n^c\rfloor\right)-\frac{1}{c}\sum_{1\le m\le N^c}\varphi(m)m^{\frac{1}{c}-1}\right|$$
$$=\left|\sum_{i\ge 1}\left(\sum_{N/2^i<n\le N/2^{i-1}}\varphi\left(\lfloor n^c\rfloor\right)-\frac{1}{c}\sum_{(N/2^i)^c<n\le(N/2^{i-1})^c}\varphi(m)m^{\frac{1}{c}-1}\right)\right|$$
$$\le\sum_{i\ge 1}F\left(\frac{N}{2^i}\right)\ll CN^{1-\rho+\varepsilon}.$$

From this the assertion follows. □

4.3.1 The Thue-Morse sequence

In our first application we are interested in the special case that the function φ is the Thue-Morse sequence in the form $\varphi(n)=(-1)^{s_2(n)}$, where $s_2(n)$ denotes the sum of digits of n in base 2.

Theorem 4.4 (The Thue-Morse sequence on $\lfloor n^c\rfloor$)**.** *There exists $a\in[0,0.4076)$ such that for all $c\in(1,2)$ and all $\eta\in\left(0,\frac{2-(a+1)c}{3-a}\right)$ there is a constant $C=C(c,\eta)$ such that for all $N\ge 2$*

$$\frac{1}{N}\left|\sum_{1\le n\le N}(-1)^{s_2(\lfloor n^c\rfloor)}\right|\le CN^{-\eta}.$$

In particular, for $1<c\le 1.42$ there exist $\eta>\max\{0,(7-5c)/9\}$ and C such that this estimate holds.

In order to prove this, we want to apply Corollary 4.3 and therefore we have to find an estimate for the expression on the left hand side of (4.9). We use the following statement which follows from Théorème 3 and inequality (1.5) in the paper [27] by Fouvry and Mauduit.

Lemma 4.5. *There exists a real number $\rho\in(0.6543,0.6632)$ such that*

$$\int_0^1\prod_{0\le k<\lambda}\left|\sin\left(2^k\pi\vartheta\right)\right|\mathrm{d}\vartheta\asymp\rho^\lambda$$

for all $\lambda\ge 0$.

The number ρ is clearly uniquely determined. No simple representation of ρ seems to be known and in fact the above bounds were obtained with the help of numerical computations. The authors of the cited article also remark that evaluating the numerical value of the integral for about a dozen values of λ (by means of splitting up the interval $[0,1]$ into 2^λ subintervals of equal length and using the fact that for $k<\lambda$ the function $\sin\left(2^k\pi\vartheta\right)$ has a constant sign on each of them) suggests that $\rho=0.661\ldots$. From Lemma 4.5 we deduce the following estimate, which is the main component of the proof of Theorem 4.4.

Proposition 4.6. *Let ρ be defined as in Lemma 4.5. Then uniformly for $z \geq 1$ we have*

$$\int_0^1 \sup_{x\geq 0} \left| \sum_{x<m\leq x+z} (-1)^{s_2(m)} \mathrm{e}\left(m\vartheta\right) \right| \mathrm{d}\vartheta \ll z^{1+\frac{\log\rho}{\log 2}}.$$

Proof. If L is an interval of the form $[\ell 2^\lambda, (\ell+1)2^\lambda)$, where ℓ and λ are nonnegative integers, we have the equality

$$\left| \sum_{m\in L} (-1)^{s_2(m)} \mathrm{e}(m\vartheta) \right| = \prod_{0\leq k<\lambda} \left|1 - \mathrm{e}\left(2^k\vartheta\right)\right|. \tag{4.13}$$

This is clear for $\lambda = 0$. If $\lambda > 0$, then by the relations $s_2(2m) = s_2(m)$ and $s_2(2m+1) = s_2(2m)+1$ we have

$$\begin{aligned}
&\left| \sum_{m\in L} (-1)^{s_2(m)} \mathrm{e}(m\vartheta) \right| \\
&\qquad = \left| \sum_{\ell 2^{\lambda-1}\leq m<(\ell+1)2^{\lambda-1}} \left((-1)^{s_2(2m)} \mathrm{e}(2m\vartheta) + (-1)^{s_2(2m+1)} \mathrm{e}((2m+1)\vartheta)\right) \right| \\
&\qquad\qquad = |(1-e(\vartheta))| \left| \sum_{\ell 2^{\lambda-1}\leq m<(\ell+1)2^{\lambda-1}} (-1)^{s_2(m)} \mathrm{e}(2m\vartheta) \right|
\end{aligned}$$

from which (4.13) follows by induction. Using the identity $|1-\mathrm{e}(\vartheta)| = 2\left|\sin(\pi\vartheta)\right|$ we get

$$\left| \sum_{m\in L} (-1)^{s_2(m)} \mathrm{e}(m\vartheta) \right| = 2^\lambda \prod_{0\leq k<\lambda} \left|\sin\left(2^k\pi\vartheta\right)\right|. \tag{4.14}$$

If L is any finite nonempty interval of nonnegative integers, we use dyadic decomposition of L in the form of the following statement: Let $a < b$ be nonnegative integers. There exists a decomposition $a = a_0 \leq \ldots \leq a_L = b_L \leq \ldots \leq b_0 = b$ such that for $j < L$ we have $a_{j+1} - a_j \in \{0, 2^j\}$, $2^j \mid a_j$ and $b_j - b_{j+1} \in \{0, 2^j\}$ and $2^j \mid b_j$.

To prove this, one first establishes the special case that $a < 2^K \leq b < 2^{K+1}$ for some K and obtains the general case by adding a multiple of 2^{K+1}. We skip the details of the proof since we will return to a very similar problem in Section 4.3.3. We can therefore decompose L into intervals of the form $[\ell 2^\lambda, (\ell+1)2^\lambda)$ in such a way that for each λ there are at most 2 such intervals of length 2^λ. From this we obtain, using (4.14), that

$$\left| \sum_{m\in L} (-1)^{s_2(m)} \mathrm{e}(m\vartheta) \right| \ll \sum_{0\leq\lambda\leq\frac{\log|L|}{\log 2}} 2^\lambda \prod_{0\leq k<\lambda} \left|\sin\left(2^k\pi\vartheta\right)\right|.$$

By Lemma 4.5 (note that in particular $2\rho > 1$) this implies

$$\int_0^1 \sup_{x\geq 0}\left|\sum_{x<m\leq x+z}(-1)^{s_2(m)}\,\mathrm{e}\left(m\vartheta\right)\right|\mathrm{d}\vartheta \ll \sum_{\lambda\leq\frac{\log(z+1)}{\log 2}} 2^\lambda \int_0^1 \prod_{k<\lambda}\left|\sin\left(2^k\pi\vartheta\right)\right|\mathrm{d}\vartheta$$
$$\ll \sum_{\lambda\leq\frac{\log(z+1)}{\log 2}} 2^\lambda\rho^\lambda \ll (2\rho)^{\frac{\log(z+1)}{\log 2}+1} \ll (2\rho)^{\frac{\log z}{\log 2}} = z^{1+\frac{\log\rho}{\log 2}}$$

for all $z\geq 1$. □

Proof of Theorem 4.4. Note first that $1+\frac{\log\rho}{\log 2}<0.4076$ according to the estimate $\rho<0.6632$. Combining Proposition 4.6 and Corollary 4.3 we get the following statement: there exists $a<0.4076$ such that for all $c\in(1,2)$ and all $\eta\in\left(0,\frac{2-(a+1)c}{3-a}\right)$ there exists C such that for all $N\geq 2$ we have

$$\frac{1}{N}\left|\sum_{1\leq n\leq N}(-1)^{s_2(\lfloor n^c\rfloor)} - \frac{1}{c}\sum_{1\leq m\leq N^c}(-1)^{s_2(m)}m^{\frac{1}{c}-1}\right| \leq CN^{-\eta}. \tag{4.15}$$

To prove the main statement, it remains to eliminate the second sum in this inequality. For all nonnegative integers K we have $\sum_{m<2K}(-1)^{s_2(m)}=0$, therefore it follows by partial summation that

$$\frac{1}{N}\sum_{1\leq m\leq N^c}(-1)^{s_2(m)}m^{\frac{1}{c}-1} \ll \frac{1}{N}\left(N^c\right)^{\frac{1}{c}-1}\sup_{1\leq u\leq N^c}\left|\sum_{1\leq m\leq u}(-1)^{s_2(m)}\right| \ll N^{-c}.$$

This quantity is dominated by the error term, so we may remove the second sum in (4.15). To finish the proof, we note that $2-(a+1)c>0$ and $\frac{7-5c}{9}<\frac{2-(a+1)c}{3-a}$ for $c\leq 1.42$ and $a<0.4076$. □

We remark that our method even yields a value around 1.425 for the upper bound on c, if indeed ρ is around 0.661 as the computations suggest. In [27, p.579], an analogous remark on the dependence of a parameter on ρ is made.

4.3.2 The joint distribution of sum-of-digits functions

For integers $q\geq 2$ and $n\geq 0$ we denote by $s_q(n)$ the sum-of-digits of n in base q. In this section we prove the following independence result of sum-of-digits functions with respect to coprime bases q_1 and q_2.

Theorem 4.7 (Joint distribution of sum-of-digits functions on $\lfloor n^c\rfloor$)**.** *Let $q_1,q_2\geq 2$, $m_1,m_2,\geq 1$ and l_1,l_2 be integers such that $(q_1,q_2)=1$, $(m_1,q_1-1)=1$ and $(m_2,q_2-1)=1$. Let $1<c<18/17$. There exists $\varepsilon>0$ such that*

$$\left|\left\{n\leq x : s_{q_1}\left(\lfloor n^c\rfloor\right)\equiv l_1 \bmod m_1 \text{ and } s_{q_2}\left(\lfloor n^c\rfloor\right)\equiv l_2 \bmod m_2\right\}\right|$$
$$= \frac{x}{m_1m_2}+O\left(x^{1-\varepsilon}\right). \tag{4.16}$$

Generalizing this theorem (and its proof) to more than two bases is straightforward, however the upper bound on c that we can obtain using our method has then to be adjusted. In order to prove Theorem 4.7, we estimate the relevant integral as well as the integrand at $\vartheta = 0$.

Proposition 4.8. *Let $q_1, q_2 \geq 2$ be relatively prime integers. There exists $C = C(q_1, q_2)$ such that for all $\alpha, \beta \in \mathbb{R}$ and $z \geq 1$ we have*

$$\int_0^1 \sup_{x \geq 0} \left| \sum_{x < n \leq x+z} \mathrm{e}\left(\alpha s_{q_1}(n) + \beta s_{q_2}(n) + n\vartheta\right) \right| \mathrm{d}\vartheta \leq C z^{8/9}. \tag{4.17}$$

Moreover, we have

$$\sup_{x \geq 0} \left| \sum_{x < n \leq x+z} \mathrm{e}\left(\alpha s_{q_1}(n) + \beta s_{q_2}(n)\right) \right| \leq C_1 z^{1-\eta(\alpha)} \tag{4.18}$$

for $z \geq 1$, where $\eta(\alpha) = \frac{\|(q_1-1)\alpha\|^2}{15 \log q_1}$ and C_1 may depend on α, β, q_1 and q_2.

In the proof of this proposition we make use of the truncated sum-of-digits function $s_{q,\lambda}$, which adds up the first λ digits of the base-q representation of a nonnegative integer n. That is, if $n = \sum_{i \geq 0} \varepsilon_i q^i$ and $\varepsilon_i \in \{0, \dots, q-1\}$ for all i, then

$$s_{q,\lambda}(n) = \sum_{0 \leq i < \lambda} \varepsilon_i = s_q\left(n \bmod q^\lambda\right).$$

For convenience we extend $s_{q,\lambda}$ to a q^λ-periodic function on $\mathbb{Z}$. By periodicity, we can represent the function $\mathrm{e}\left(\alpha s_{q,\lambda}(n)\right)$ with the aid of the discrete Fourier transform. For integers $q \geq 2$, $\lambda \geq 0$ and n we have

$$\mathrm{e}\left(\alpha s_{q,\lambda}(n)\right) = \sum_{h < q^\lambda} \mathrm{e}\left(hnq^{-\lambda}\right) F_{q,\lambda}(h, \alpha) \tag{4.19}$$

and

$$\mathrm{e}\left(-\alpha s_{q,\lambda}(n)\right) = \sum_{h < q^\lambda} \mathrm{e}\left(hnq^{-\lambda}\right) \overline{F_{q,\lambda}(-h, \alpha)}, \tag{4.20}$$

where

$$F_{q,\lambda}(h, \alpha) = \frac{1}{q^\lambda} \sum_{u < q^\lambda} \mathrm{e}\left(\alpha s_{q,\lambda}(u) - huq^{-\lambda}\right).$$

The Fourier coefficients $F_{q,\lambda}(h, \alpha)$ may be estimated uniformly in h using the following lemma ([41, Lemme 9]).

Lemma 4.9. *Let $q, \lambda \geq 2$ and h be integers and $\alpha \in \mathbb{R}$. Then*

$$|F_{q,\lambda}(h, \alpha)| \leq \mathrm{e}^{\pi^2/48}\, q^{-c_q \|(q-1)\alpha\|^2 \lambda},$$

where

$$c_q = \frac{\pi^2}{12 \log q} \left(1 - \frac{2}{q+1}\right).$$

We prove the following lemma on the truncated sum-of-digits function, which is a way of expressing the idea that addition of an integer r to n should only change digits at low positions in most cases.

Lemma 4.10. *Let $q \geq 2$, $\lambda \geq 0$ and r be integers and let I be a finite interval in $\mathbb{N}$ such that $I + r \subseteq \mathbb{N}$. Then*

$$|\{n \in I : s_q(n+r) - s_q(n) \neq s_{q,\lambda}(n+r) - s_{q,\lambda}(n)\}| \leq |I| \, \frac{|r|}{q^\lambda} + |r| \, .$$

Proof. It is sufficient to assume that r is nonnegative, since the other case then follows by shifting the interval I.

For a nonnegative integer n, there exist unique t and u such that $n = tq^\lambda + u$, where $u < q^\lambda$. Clearly we have $s_q(n) = s_q(t) + s_q(u)$ and $s_{q,\lambda}(n) = s_q(u)$. If $n \equiv k \bmod q^\lambda$ for some k such that $0 \leq k < q^\lambda - r$, then $s_q(n+r) = s_q(t) + s_q(u+r)$ and $s_{q,\lambda}(n+r) = s_q(u+r)$, therefore $s_q(n+r) - s_q(s) = s_{q,\lambda}(n+r) - s_{q,\lambda}(n)$. It remains therefore to show that $\big|\{n \in I : q^\lambda - r \leq n \bmod q^\lambda < q^\lambda\}\big| \leq |I| \, r/q^\lambda + r$, which is not difficult. □

The inequality of van der Corput is well known. For our purposes, we will employ it in the following form.

Lemma 4.11. *Let I be a finite interval in $\mathbb{Z}$ and let $a_n \in \mathbb{C}$ for $n \in I$. Then*

$$\left| \sum_{n \in I} a_n \right|^2 \leq \frac{|I| - 1 + R}{R} \sum_{0 \leq |r| < R} \left(1 - \frac{|r|}{R}\right) \sum_{\substack{n \in I \\ n+r \in I}} a_{n+r} \overline{a_n}$$

for all integers $R \geq 1$.

Proof of Proposition 4.8. To estimate the left hand side of (4.17), we introduce two parameters to be chosen later, λ_1 and λ_2. Rounding off z to the nearest multiple M of $q_1^{\lambda_1} q_2^{\lambda_2}$ introduces an error term $O\left(q_1^{\lambda_1} q_2^{\lambda_2}\right)$. Let $x \geq 0$, $z \geq 1$ and let $R \in [1, z]$ be an integer. Then by van der Corput's inequality we get

$$\begin{aligned} \left| \sum_{x < n \leq x+M} \mathrm{e}\left(\alpha s_{q_1}(n) + \beta s_{q_2}(n) + n\vartheta\right) \right|^2 &\ll \frac{z}{R} \sum_{|r| < R} \left(1 - \frac{|r|}{R}\right) \\ &\times \sum_{x < n, n+r \leq x+M} \mathrm{e}\big(\alpha \left(s_{q_1}(n+r) - s_{q_1}(n)\right) + \beta \left(s_{q_2}(n+r) - s_{q_2}(n)\right) + r\vartheta\big). \end{aligned}$$

Applying Lemma 4.10 in order to replace s_{q_1} and s_{q_2} by s_{q_1,λ_1} and s_{q_2,λ_2} respectively and omitting the summation condition $x < n + r \leq x + M$ afterwards we get an error term $O\left(zR + z^2 R\left(1/q_1^{\lambda_1} + 1/q_2^{\lambda_2}\right)\right)$ and after inserting equations (4.19) and (4.20) it remains to estimate the quantity

$$\frac{z}{R^2} \sum_{\substack{h_1, k_1 < q_1^{\lambda_1} \\ h_2, k_2 < q_2^{\lambda_2}}} F_{q_1,\lambda_1}(h_1, \alpha) \overline{F_{q_1,\lambda_1}(-k_1, \alpha)} F_{q_2,\lambda_2}(h_2, \beta) \overline{F_{q_2,\lambda_2}(-k_2, \beta)}$$

$$\times \sum_{x<n\le x+M} \mathrm{e}\left(n\left(\frac{h_1+k_1}{q_1^{\lambda_1}}+\frac{h_2+k_2}{q_2^{\lambda_2}}\right)\right) \sum_{|r|<R} (R-|r|)\,\mathrm{e}\left(r\left(\frac{h_1}{q_1^{\lambda_1}}+\frac{h_2}{q_2^{\lambda_2}}+\vartheta\right)\right). \quad (4.21)$$

By our choice of M and by the Chinese Remainder Theorem, the contribution of the case that $(h_1+k_1, h_2+k_2) \not\equiv (0,0) \bmod \left(q_1^{\lambda_1}, q_2^{\lambda_2}\right)$ is 0. Using the identity

$$\sum_{|r|<R} (R-|r|)\,\mathrm{e}(rx) = \left|\sum_{r<R} \mathrm{e}(rx)\right|^2,$$

we see that (4.21) is bounded by the expression

$$\frac{z^2}{R^2} \sum_{\substack{h_1<q_1^{\lambda_1}\\ h_2<q_2^{\lambda_2}}} |F_{q_1,\lambda_1}(h_1,\alpha)|^2\, |F_{q_2,\lambda_2}(h_2,\beta)|^2 \left|\sum_{|r|<R} \mathrm{e}\left(r\left(\frac{h_1}{q_1^{\lambda_1}}+\frac{h_2}{q_2^{\lambda_2}}+\vartheta\right)\right)\right|^2, \quad (4.22)$$

which is independent of x. In order to prove the first part of Proposition 4.8, we use the Cauchy-Schwarz inequality, Parseval's identity and the identity

$$\int_0^1 \left|\sum_{r\in I} \mathrm{e}\left(r(t+\vartheta)\right)\right|^2 \mathrm{d}\vartheta = |I|$$

and collect the error terms to arrive at the estimate

$$\int_0^1 \sup_{x\ge 0} \left|\sum_{x<n\le x+z} \mathrm{e}\left(\alpha s_{q_1}(n)+\beta s_{q_2}(n)+n\vartheta\right)\right| \mathrm{d}\vartheta$$
$$= O\left(q_1^{\lambda_1} q_2^{\lambda_2} + z^{1/2}R^{1/2} + zR^{1/2}\left(q_1^{-\lambda_1/2}+q_2^{-\lambda_2/2}\right) + zR^{-1/2}\right), \quad (4.23)$$

which is valid for all real α, β and $z \ge 1$ and all integers $R \in [1,z]$ and $\lambda_1, \lambda_2 \ge 0$. The implied constant is an absolute one. This estimate is also valid for real R, λ_1 and λ_2, however the implied constant may then depend on q_1 and q_2. We set

$$\lambda_1 = \frac{4\log z}{9\log q_1}, \quad \lambda_2 = \frac{4\log z}{9\log q_2} \quad \text{and} \quad R = z^{2/9}.$$

Then clearly $R \in [1,z]$ and a short calculation shows that all four summands in the error term are $\ll z^{8/9}$, which proves the first part. For the second part we make use of Lemma 4.9 and Parseval's identity to estimate (4.22) by

$$\frac{z^2}{R^2} \sup_{h\in\mathbb{Z}} |F_{q_1,\lambda_1}(h,\alpha)|^2 \sup_{t\in\mathbb{R}} \left|\sum_{h_1<q_1^{\lambda_1}} \min\left\{R^2, \left\|h_1/q_1^{\lambda_1}+t\right\|^{-2}\right\}\right|$$
$$\times \sum_{h_2<q_2^{\lambda_2}} |F_{q_2,\lambda_2}(h_2,\beta)|^2 \ll z^2 q_1^{-2c\lambda_1} \frac{q_1^{\lambda_1}}{R}, \quad (4.24)$$

where $c = c_{q_1} \|(q_1-1)\alpha\|^2$. Therefore for some constant C the following holds for all $x, z \geq 0$ and all integers $R \in [1, z]$.

$$\left| \sum_{x<n\leq x+z} \mathrm{e}\left(\alpha s_{q_1}(n) + \beta s_{q_2}(n)\right) \right| \leq C\left(q_1^{\lambda_1} q_2^{\lambda_2} + z^{1/2}R^{1/2} + zR^{1/2}\left(q_1^{-\lambda_1/2} + q_2^{-\lambda_2/2}\right) + zq_1^{\lambda_1(1/2-c)}R^{-1/2}\right).$$

Again we may assume that R, λ_1 and λ_2 are real numbers. We set

$$\lambda_1 = \frac{2\log z}{(4+c)\log q_1}, \quad \lambda_2 = \frac{2\log z}{(4+c)\log q_2} \text{ and } \quad R = z^{\frac{2-2c}{4+c}}.$$

With these choices we get after a short calculation

$$\sum_{x<n\leq x+z} \mathrm{e}\left(\alpha s_{q_1}(n) + \beta s_{q_2}(n)\right) \ll z^{1-c/(4+c)}.$$

To get a convenient form of the exponent, we note that $q_1 \geq 2$, which implies $c_{q_1} \geq \pi^2/(36\log q_1)$. By the same condition and monotonicity of $x/(4+x)$ we get

$$\frac{c}{4+c} \geq \frac{\pi^2\|(q_1-1)\alpha\|^2}{36\log q_1\left(4+\frac{\pi^2\|(q_1-1)\alpha\|^2}{36\log q_1}\right)} \geq \frac{\|(q_1-1)\alpha\|^2}{\frac{144\log q_1}{\pi^2}+\frac{1}{4}} \geq \frac{\|(q_1-1)\alpha\|^2}{15\log q_1}.$$

□

By Corollary 4.3 and (4.17) we see that for all real α and β the function $\varphi(m) = \mathrm{e}\left(\alpha s_{q_1}(m) + \beta s_{q_2}(m)\right)$ admits a "change of variables" as long as $2-(8/9+1)c > 0$, that is, $c < 18/17$. We assume now that $(q_1-1)\alpha \notin \mathbb{Z}$ or $(q_2-1)\beta \notin \mathbb{Z}$. Then by partial summation and equation (4.18) the second sum in (4.10) can be eliminated, leading to the following statement:

Let $q_1, q_2 \geq 2$ be relatively prime and $\alpha, \beta \in \mathbb{R}$ such that $(q_1-1)\alpha \notin \mathbb{Z}$ or $(q_2-1)\beta \notin \mathbb{Z}$. Then for all $c \in (1, 18/17)$ there exist $\varepsilon > 0$ and C such that for $N \geq 1$ we have

$$\sum_{1\leq n\leq N} \mathrm{e}\left(\alpha s_{q_1}\left(\lfloor n^c \rfloor\right) + \beta s_{q_2}\left(\lfloor n^c \rfloor\right)\right) \leq CN^{1-\varepsilon}.$$

From this exponential sum estimate we get the statement of Theorem 4.7 by an orthogonality argument, which completes the proof.

Note that by the same orthogonality argument (4.4) can be deduced from from (4.18), which gives an alternative to Kim's proof [33].

4.3.3 The Zeckendorf sum-of-digits function

In our third application we study the distribution in residue classes of the values of the Zeckendorf sum-of-digits function on $\lfloor n^c \rfloor$.

For $k \geq 0$ let F_k be the k-th Fibonacci number, that is, $F_0 = 0$, $F_1 = 1$ and $F_k = F_{k-1} + F_{k-2}$ for $k \geq 2$. By Zeckendorf's Theorem [57] every positive integer n admits a unique representation

$$n = \sum_{i \geq 2} \varepsilon_i F_i,$$

where $\varepsilon_i \in \{0, 1\}$ and $\varepsilon_i = 1 \Rightarrow \varepsilon_{i+1} = 0$. By this theorem we may write the i-th coefficient ε_i as a function of n. The Zeckendorf sum-of-digits of n is then defined as

$$s_Z(n) = \sum_{i \geq 2} \varepsilon_i(n).$$

We set $s_Z(0) = 0$. We note that $s_Z(n)$ is the least k such that n is the sum of k Fibonacci numbers.

Theorem 4.12 (The Zeckendorf sum-of-digits function on $\lfloor n^c \rfloor$)**.** *Let $m \geq 1$ and a be integers. Then for all $c \in (1, 4/3)$ there exists $\varepsilon > 0$ such that uniformly for $x \geq 1$ we have*

$$|\{n \leq x : s_Z(\lfloor n^c \rfloor) \equiv a \bmod m\}| = \frac{x}{m} + O\left(x^{1-\varepsilon}\right).$$

The proof of this statement is based on the following proposition.

Proposition 4.13. *There exist C such that for all $\alpha \in \mathbb{R}$ and $z \geq 1$ we have*

$$\int_0^1 \sup_{x \geq 0} \left| \sum_{x < n \leq x+z} \mathrm{e}\left(\alpha s_Z(n) + n\vartheta\right) \right| \mathrm{d}\vartheta \leq C z^{1/2}. \tag{4.25}$$

Moreover for $\alpha \notin \mathbb{Z}$ there exist $\eta > 0$ and C_1 such that for all $z \geq 1$

$$\sup_{x \geq 0} \left| \sum_{x < n \leq x+z} \mathrm{e}\left(\alpha s_Z(n)\right) \right| \leq C_1 z^{1-\eta}. \tag{4.26}$$

Proof. For $k \geq 0$ we define

$$G_k(\alpha, \vartheta) = \sum_{0 \leq u < F_k} \mathrm{e}\left(\alpha s_Z(u) + \vartheta u\right).$$

By the Cauchy-Schwarz inequality and the formula $F_k \asymp \varphi^k$, where $\varphi = (\sqrt{5}+1)/2$, we clearly have

$$\int_0^1 \left| \sum_{n < F_k} G_k(\alpha, \vartheta) \right| \mathrm{d}\vartheta \leq F_k^{1/2} \ll \varphi^{k/2}. \tag{4.27}$$

Moreover, by the relation $s_Z(u + F_k) = 1 + s_Z(u)$ that holds for $k \geq 2$ and $0 \leq u < F_{k-1}$ the terms $G_k(\alpha, 0)$ satisfy the linear recurrence relation

$$G_{k+1}(\alpha, 0) = G_k(\alpha, 0) + \mathrm{e}(\alpha) G_{k-1}(\alpha, 0).$$

Its characteristic polynomial has the roots $\frac{1}{2} \pm \frac{1}{2}\sqrt{1+4\,\mathrm{e}(\alpha)}$, whose absolute values are bounded by $\frac{1}{2}+\frac{1}{2}(17+8\cos(2\pi\alpha))^{1/4}$. This expression is equal to φ if $\alpha \in \mathbb{Z}$ and strictly less than φ otherwise. Consequently, if $\alpha \notin \mathbb{Z}$, there is some $\eta > 0$ such that

$$G_k(\alpha, 0) \ll \varphi^{k(1-\eta)}. \tag{4.28}$$

The expression for $G_k(\alpha, \vartheta)$ involves a sum over the interval $[0, F_k)$. In order to deal with arbitrary finite intervals I in $\mathbb{N}$, we decompose the interval I according to the Zeckendorf representation of its endpoints. This procedure is analogous to the decomposition of an interval into dyadic intervals, which we used in the proof of Theorem 4.4.

Lemma 4.14. *Let $0 \le A < B$ be integers. There exist integers $L \ge 2$ and a_j, b_j for $2 \le j \le L$ such that $A = a_2 \le \cdots \le a_L = b_L \le \cdots \le b_2 = B$ having the properties that $\varepsilon_i(a_j) = \varepsilon_i(b_j) = 0$ for $2 \le i < j \le L$ and that $a_{j+1} - a_j \in \{0, F_{j-1}\}$ and $b_j - b_{j+1} \in \{0, F_j\}$ for $2 \le j < L$.*

Proof. We first show that it is sufficient to assume that $0 \le A < F_K \le B < F_{K+1}$ for some $K \ge 2$. Let $K = \max\{i : \varepsilon_i(A) \ne \varepsilon_i(B)\}$ and $C = \sum_{i>K} \varepsilon_i(A)F_i = \sum_{i>K} \varepsilon_i(B)F_i$. Then $0 \le A - C < F_K \le B - C < F_{K+1}$ and by our assumption we get a decomposition $A-C = a_2 \le \cdots \le a_L = b_L \le \cdots \le b_2 = B-C$ as in the Lemma. We have $\varepsilon_i(a_j) = \varepsilon_i(b_j) = 0$ for $2 \le j \le L$ and $i > K$ and since $\varepsilon_K(B) = 1$, we have $\varepsilon_i(C) = 0$ for $i \le K+1$. Therefore $A = a_2 + C \le \cdots \le a_L + C = b_L + C \le \cdots \le b_2 + C = B$ is a valid decomposition of the interval $[A, B]$.

It remains to prove the simplified statement. In the case that $A = 0$ we set $a_2 = \ldots = a_{K+1} = 0$ and $b_j = \sum_{i \ge j} \varepsilon_i(B)F_i$ for $2 \le j \le K+1$. Otherwise we set $b_j = \sum_{i \ge j} \varepsilon_i(B)F_i$ for $2 \le j \le K$ and to choose a_j, we use the following assertion which we prove by (downward) induction on k.

> Let $K \ge 2$. Assume that $0 < A \le F_K$ and $k = \min\{i : \varepsilon_i(A) = 1\}$. There exist integers $A = a_k \le \cdots \le a_K = F_K$ such that for $k \le j < K$ and $2 \le i < j$ we have $\varepsilon_i(a_j) = 0$ and $a_{j+1} - a_j \in \{0, F_{j-1}\}$.

If $k = K$, then $A = F_K$ and we choose $a_K = A$. Otherwise $2 \le k < K$ and we set $A' = A + F_{k-1}$ and $k' = \min\{i : \varepsilon_i(A') = 1\}$. Then $k' > k$. We choose $a_{k'}, \ldots, a_K$ according to the assumption, $a_k = A$ and $a_{k+1} = \cdots = a_{k'-1} = A'$. This choice gives an admissible decomposition of the interval $[A, F_K]$ and the statement is proved. Setting $a_2 = \cdots = a_{k-1} = A$ completes the proof of Lemma 4.14. □

By this lemma we can decompose an arbitrary finite interval in $\mathbb{N}$ into intervals of the form $[A, A + F_j)$, where $\varepsilon_i(A) = 0$ for $i \le j$, in such a way that for each $j \ge 1$ there are at most 2 intervals of this form. Noting also that $s_Z(n) = s_Z(A) + s_Z(n-A)$ for all n in such an interval and using the formula $F_k \asymp \varphi^k$, one can easily derive (4.25) and (4.26) from (4.27) and (4.28). □

We plug (4.25) into Corollary 4.3 and eliminate the second sum in (4.10) by partial summation and (4.26), which results in the statement that for $\alpha \in \mathbb{R} \setminus \mathbb{Z}$ and for $c \in (1, 4/3)$ there exist $\eta > 0$ and C such that

$$\sum_{1 \le n \le N} \mathrm{e}\left(\alpha s_Z\left(\lfloor n^c \rfloor\right)\right) \le C N^{1-\eta}$$

for $N \geq 1$. By transferring this to a statement about residue classes, we obtain the statement of Theorem 4.12.

4.4 Proofs of the main results

We start with a couple of lemmas that we need in the proofs of Theorem 4.1 and Proposition 4.2. The first one will allow proving that the left hand sides of (4.5) and (4.7) are always $O(A)$.

Lemma 4.15. *Let $f : \mathbb{R}^+ \to \mathbb{R}^+$ be differentiable and assume that f' is increasing and positive. Then*

$$\sum_{f(A)<m\leq f(2A)} \left(f^{-1}\right)'(m) \ll A$$

for $A > 0$.

Proof. If $g : \mathbb{R}^+ \to \mathbb{R}^+$ is decreasing and $0 < s \leq t$, we have

$$\sum_{s<m\leq t} g(m) = \sum_{\lfloor s\rfloor+1\leq m\leq \lfloor t\rfloor} g(m) = \int_{\lfloor s\rfloor}^{\lfloor t\rfloor} g\left(\lfloor x\rfloor + 1\right) \mathrm{d}x$$
$$\leq g\left(\lfloor s\rfloor + 1\right) + \int_{\lfloor s\rfloor+1}^{\lfloor t\rfloor} g(x)\,\mathrm{d}x \leq g(s) + \int_s^t g(x)\,\mathrm{d}x.$$

We apply this to the function $g(x) = \left(f^{-1}\right)'(x)$, noting also that there is some $a > 0$ such that the sum in the lemma is equal to 0 for $A < a$. For $A \geq a$ we have

$$\sum_{f(A)<m\leq f(2A)} \left(f^{-1}\right)'(m) \leq \frac{1}{f'(A)} + f^{-1}(x)\Big|_{f(A)}^{f(2A)} \leq \frac{1}{f'(a)} + A \ll A.$$

□

In the next lemma we study properties of functions f as in Theorem 4.1 and Proposition 4.2.

Lemma 4.16. *Assume that $f : \mathbb{R}^+ \to \mathbb{R}$ is two times continuously differentiable, $f, f', f'' > 0$ and that there exist $c_1 \geq 1/2$ and $c_2 > 0$ such that for $0 < x \leq y \leq 2x$ we have $c_1 f''(x) \leq f''(y) \leq c_2 f''(x)$. Then the following estimates hold.*

$$xf''(x) \ll yf''(y) \qquad \text{for } 0 < x \leq y \tag{4.29}$$
$$xf''(x) \ll f'(x) \ll xf''(x)\log x \qquad \text{for } x \geq 2 \tag{4.30}$$
$$f'(x) \leq f'(y) \ll f'(x) \qquad \text{for } 0 < x \leq y \leq 2x, \tag{4.31}$$
$$\log x \ll f'(x) \ll x^{\delta} \qquad \text{for some } \delta \geq 0 \text{ and all } x \geq 2. \tag{4.32}$$

Moreover for $0 < x \leq a \leq b \leq 2x$ we have

$$f(b) - f(a) \asymp f'(x)(b-a) \tag{4.33}$$

and

$$f'(b) - f'(a) \asymp f''(x)(b-a). \tag{4.34}$$

Proof. In order to prove (4.29), we show the equivalent statement that

$$f''(x) \ll af''(ax)$$

for $a \geq 1$ and $x > 0$. This is clear for $a = 2^k$ by the inequalities $c_1 f''(x) \leq f''(2x)$ and $c_1 \geq 1/2$. If $2^k \leq a < 2^{k+1}$, we have $f''(ax) \geq c_1 f''(2^k x) \geq c_1 2^{-k} f''(x) \gg 1/af''(x)$. We turn to the first inequality in (4.30). By the Mean Value Theorem there exists some $\xi \in [x/2, x]$ such that $f'(x) \geq f'(x) - f'(x/2) = x/2f''(\xi) \geq x/(2c_2)f''(x)$. For the proof of the second inequality in (4.30), let $x \geq 2$. For $t \leq x$ we have $tf''(t) \ll xf''(x)$ by (4.29) and therefore

$$f'(x) = f'(2) + \int_2^x f''(t)\,\mathrm{d}t \ll f'(2) + xf''(x)\int_2^x \frac{1}{t}\,\mathrm{d}t \leq f'(2) + xf''(x)\log x.$$

For $x \geq 2$ we have $xf''(x)\log x \gg f''(2) \gg f'(2)$ by (4.29) and $f', f'' > 0$, therefore $f'(x) \ll xf''(x)\log x$. The first inequality of (4.31) is obvious since f' is increasing. By applying the Mean Value Theorem it follows that there exists $\xi \in [x, 2x]$ such that $f'(2x) - f'(x) = xf''(\xi) \ll xf''(x)$. Together with (4.30) we get $f'(2x) \ll f'(x)$. We prove (4.32). The first estimate follows from (4.29) if we set $x = 1$ and integrate in y. By (4.31) there exists $c > 0$ such that $f'(2z) \leq cf'(z)$ for all $z > 0$, from which we get $f'(x) \ll c^{\frac{\log x}{\log 2}} f'(1)$ for all $x \geq 1$. Let $0 < x \leq a \leq b \leq 2x$. By the Mean Value Theorem there is some $\xi \in [a, b]$ such that $f(b) - f(a) = f'(\xi)(b - a)$. From the monotonicity of f' and (4.31) we get (4.33). Analogously, (4.34) is proved via the assumption $c_2 f''(x) \leq f''(y) \leq c_2 f''(x)$. □

In the following lemma we integrate over a well-known estimate for the exponential sum $\sum \mathrm{e}(nx)$, where the sum extends over an interval containing B integers.

Lemma 4.17. *Let $a \leq b$ be real numbers and $B \geq 2$. Then*

$$\int_a^b \min\left\{B, \|x\|^{-1}\right\}\mathrm{d}x \leq 2\,(b - a + 1)\,(1 + \log B)\,.$$

Proof. Since the integrand is 1-periodic and symmetric with respect to $\frac{1}{2}$, we have

$$\begin{aligned}\int_a^b \min\left\{B, \|x\|^{-1}\right\}\mathrm{d}x &\leq 2(b - a + 1)\int_0^{1/2} \min\left\{B, \|x\|^{-1}\right\}\mathrm{d}x \\ &\leq 2(b - a + 1)\left(\int_0^{1/B} B\,\mathrm{d}x + \int_{1/B}^{1/2} x^{-1}\,\mathrm{d}x\right) \\ &\leq 2(b - a + 1)\,(1 + \log(1/2) - \log(1/B)) \leq 2\,(b - a + 1)\,(1 + \log B)\,.\end{aligned}$$

□

4.4.1 Proof of Proposition 4.2

We prepare for the proof of Proposition 4.2 by giving some results on the approximation of a twice differentiable function by an affine linear function.

Lemma 4.18. *Let $f : [a, b] \to \mathbb{R}$ be twice differentiable and $|f''| \le M$. For all $\alpha \in f'([a, b])$ and $a \le x \le b$ we have*

$$|x\alpha + f(a) - a\alpha - f(x)| \le M(b-a)^2.$$

Proof. By the Mean Value Theorem there exists some $\xi_1 \in [a, x]$ such that $f(x) - f(a) = f'(\xi_1)(x-a)$, that is, such that $|x\alpha + f(a) - a\alpha - f(x)| = (x-a)\,|f'(\xi_1) - \alpha|$. There exists some $y \in [a, b]$ such that $\alpha = f'(y)$. By applying the Mean Value Theorem to the function f', we get some ξ_2 between ξ_1 and y such that $|f'(\xi_1) - \alpha| = |f'(\xi_1) - f'(y)| = |(\xi_1 - y)f''(\xi_2)|$. From this the statement follows easily. □

The following result will permit us to replace the function $\lfloor f(n) \rfloor$ by a Beatty sequence on an interval $(a, b]$.

Lemma 4.19. *Let $f : [a, b] \to \mathbb{R}$ be twice differentiable and $|f''| \le M$. For all $\alpha \in f'([a, b])$ and $a \le x \le b$ such that $\|x\alpha + f(a) - a\alpha\| > M(b-a)^2$ we have*

$$\lfloor f(x) \rfloor = \lfloor x\alpha + f(a) - a\alpha \rfloor .$$

Proof. We write $\beta = f(a) - a\alpha$ and $d = M(b-a)^2$. The condition $\|x\alpha + \beta\| > d$ in the statement of the lemma implies $\lfloor x\alpha + \beta - d \rfloor = \lfloor x\alpha + \beta \rfloor = \lfloor x\alpha + \beta + d \rfloor$. Moreover by Lemma 4.18 we get $x\alpha + \beta - d \le f(x) \le x\alpha + \beta + d$. Combining these observations yields the claim. □

We estimate the number of integers in an interval for which such an approximation fails.

Lemma 4.20. *Let $a \le b$ be integers and let $f : [a, b] \to \mathbb{R}$ be twice differentiable. Assume that $|f''| \le M$. For all $\alpha \in f'([a, b])$ and all $R \ge 1$ we have the estimate*

$$|\{n \in (a, b] : \lfloor f(n) \rfloor \ne \lfloor n\alpha + f(a) - a\alpha \rfloor\}| \le 2M(b-a)^3 + \frac{(b-a)}{R} + \sum_{1 \le r \le R} \frac{1}{r} \left| \sum_{a < n \le b} \mathrm{e}\left(nr\alpha\right) \right| .$$

Proof. Write $d = M(b-a)^2$ and $\beta = f(a) - a\alpha$. If $d \ge \frac{1}{2}$ or $a = b$ the statement follows immediately since the left hand side is bounded by $b - a$. Otherwise it suffices by Lemma 4.19 to estimate the quantity

$$|\{n \in (a, b] : \|n\alpha + \beta\| \le d\}| .$$

To do this, we apply the inequality of Erdős and Turán to the sequence $(\{n\alpha + \beta + d\})_{a < n \le b}$ in $[0, 1)$. According to [43, Lemma 1], the discrepancy of any real valued finite sequence $(x_1, \ldots, x_N)$ in $[0, 1)$, where $N \ge 1$, satisfies

$$D_N(x_1, \ldots, x_N) = \sup_{0 \le r \le s < 1} \left| \frac{1}{N} |\{1 \le n \le N : r \le x_n \le s\}| - (s - r) \right| \le \frac{1}{H+1} + \sum_{1 \le h \le H} \frac{1}{h} \left| \frac{1}{N} \sum_{1 \le n \le N} \mathrm{e}(hx_n) \right|$$

for all $H \geq 1$. This is the classical inequality of Erdős and Turán with an improved constant, equal to 1.

Considering the interval $[0, 2d]$, we obtain from this the estimate

$$\begin{aligned}\left|\frac{1}{b-a}\left|\{n \in (a,b] : \|n\alpha+\beta\| \leq d\}\right| - 2d\right| \\ = \left|\frac{1}{b-a}\left|\{n \in (a,b] : \{n\alpha+\beta+d\} \in [0,2d]\}\right| - 2d\right| \\ \leq \frac{1}{R} + \frac{1}{b-a}\sum_{1\leq r\leq R}\frac{1}{r}\left|\sum_{a<n\leq b}\mathrm{e}\left(nr\alpha+r\beta+rd\right)\right|,\end{aligned}$$

from which the claim follows. □

The rough idea of the proof of Proposition 4.2 is to relate the two sums in (4.7) to each other in three steps, introducing the expression (4.8). We replace the function $\lfloor f(n)\rfloor$ by a Beatty sequence $\lfloor n\alpha+\beta\rfloor$ on small subintervals of $(A, 2A]$. Analogously, we replace the expression $\left(f^{-1}\right)'(m)$ by the constant value $\frac{1}{\alpha}$ on corresponding subintervals of $(f(A), f(2A)]$. To link the two expressions thus obtained we insert (4.8), which expresses the error that arises when we replace the sum of $\varphi(n)$ over a Beatty sequence by a sum of $\varphi(n)$ over all integers in an interval. Afterwards we collect the error terms and we are done.

Proof of Proposition 4.2. Let $A \geq 2$. It is sufficient to concentrate on the case that K is an integer and $2 \leq K \leq A$, for the following reasons. If $K < 2$, then $\frac{(\log A)^2}{K} \gg 1$, and if $K > A$, then $f''(A)K^2 \geq Af''(A)A \gg 2f''(2) \gg 1$ by (4.29). Therefore the right hand side of (4.7) is bounded below for these cases, while the left hand side of (4.7) is always bounded above by Lemma 4.15. For general K in $[2, A]$ we have $|I(A, \lfloor K\rfloor) - I(A, K)| \ll \frac{1}{K}$, which can be deduced from the inequality $|ab - a'b'| \leq |a-a'|\,|b| + |a'|\,|b-b'|$ and the estimate $\alpha \geq f'(2) \gg 1$ that is valid for $\alpha \in [f'(A), f'(2A)]$. This error is absorbed by the term $\frac{(\log A)^2}{K}$, therefore the general case can easily be accounted for by adjusting the implied constant C.

To guarantee that all expressions involving φ are well-defined, we set $\varphi(n) = 0$ for $n \leq 0$. For K an integer and $2 \leq K \leq A$ we partition the interval $(A, 2A]$ into smaller intervals of length at most K as follows. Define integral partition points $a_i = \lceil A\rceil + iK$ for $i \geq 0$ and set $L = \max\{i : a_i \leq 2A\}$, which is well defined since $K > 0$. The integer L satisfies the estimate $L \leq \frac{A}{K}$. We have the decomposition

$$(A, 2A] = (A, \lceil A\rceil] \cup \bigcup_{0\leq i<L}(a_i, a_{i+1}] \cup (a_L, 2A]. \tag{4.35}$$

Let $\alpha \in \mathbb{R}$. Then by the triangle inequality and the relation $a_{i+1} - a_i = K$ we have for $i < L$

$$\begin{aligned}\left|\sum_{a_i<n\leq a_{i+1}}\varphi\left(\lfloor f(n)\rfloor\right) - \sum_{f(a_i)<m\leq f(a_{i+1})}\varphi(m)\left(f^{-1}\right)'(m)\right| \\ \leq T_1(\alpha,i) + T_2(\alpha,i) + T_3(\alpha,i) + T_4(\alpha,i),\end{aligned} \tag{4.36}$$

where

$$T_1(\alpha,i)=\left|\sum_{a_i<n\le a_{i+1}}\left(\varphi\left(\lfloor f(n)\rfloor\right)-\varphi\left(\lfloor n\alpha+f(a_i)-a_i\alpha\rfloor\right)\right)\right|,$$

$$T_2(\alpha,i)=\left|\sum_{0<n\le K}\varphi\left(\lfloor n\alpha+f(a_i)\rfloor\right)-\frac{1}{\alpha}\sum_{f(a_i)<m\le f(a_i)+K\alpha}\varphi(m)\right|,$$

$$T_3(\alpha,i)=\left|\frac{1}{\alpha}\sum_{f(a_i)<m\le a_{i+1}\alpha+f(a_i)-a_i\alpha}\varphi(m)-\frac{1}{\alpha}\sum_{f(a_i)<m\le f(a_{i+1})}\varphi(m)\right|,$$

$$T_4(\alpha,i)=\left|\sum_{f(a_i)<m\le f(a_{i+1})}\varphi(m)\left(\frac{1}{\alpha}-\left(f^{-1}\right)'(m)\right)\right|.$$

We integrate (4.36) in α from $f'(a_i)$ to $f'(a_{i+1})$, divide by the length of the integration range, and take the sum over i from 0 to $L-1$, obtaining

$$\left|\sum_{\lceil A\rceil<n\le a_L}\varphi\left(\lfloor f(n)\rfloor\right)-\sum_{f(\lceil A\rceil)<m\le f(a_L)}\varphi(m)\left(f^{-1}\right)'(m)\right|$$
$$\le\sum_{0\le i<L}\frac{1}{f'(a_{i+1})-f'(a_i)}\int_{f'(a_i)}^{f'(a_{i+1})}\Big(T_1(\alpha,i)+T_2(\alpha,i)$$
$$+T_3(\alpha,i)+T_4(\alpha,i)\Big)\,\mathrm{d}\alpha. \quad (4.37)$$

The first summand will be estimated with the help of Lemma 4.20, the second by $A\,I(A,K)$, and the third and fourth terms will be estimated trivially.

We estimate the first summand in (4.37). If R is a positive integer, $0\le i<L$ and $\alpha\in f'([a_i,a_{i+1}])$, Lemma 4.20 gives

$$T_1(\alpha,i)\le 2f''(A)K^3+\frac{K}{R}+\sum_{1\le r\le R}\frac{1}{r}\left|\sum_{a_i<n\le a_{i+1}}\mathrm{e}(nr\alpha)\right|. \quad (4.38)$$

By (4.34) we have $f'(2A)-f'(A)\ll Af''(A)$ and $f'(a_{i+1})-f'(a_i)\gg f''(A)K$ for $0\le i<L$. Note also that $Af''(A)\gg 2f''(2)>0$ for all $A\ge 2$ by (4.29) and $f''>0$. From Lemma 4.17 it follows that for $2\le K\le A$ and $r\ge 1$ we have

$$\sum_{0\le i<L}\frac{1}{f'(a_{i+1})-f'(a_i)}\int_{f'(a_i)}^{f'(a_{i+1})}\left|\sum_{a_i<n\le a_{i+1}}\mathrm{e}(nr\alpha)\right|\mathrm{d}\alpha$$
$$\ll\frac{1}{f''(A)K}\sum_{0\le i<L}\frac{1}{r}\int_{rf'(a_i)}^{rf'(a_{i+1})}\left|\sum_{a_i<n\le a_{i+1}}\mathrm{e}(xn)\right|\mathrm{d}x$$

$$\leq \frac{1}{f''(A)K}\frac{1}{r}\int_{rf'(A)}^{rf'(2A)} \min\left\{K, \|x\|^{-1}\right\} \mathrm{d}x$$
$$\ll \frac{1}{f''(A)K}\frac{1}{r} 2(rAf''(A)+1)\,(1+\log K) \ll A\frac{\log K}{K}. \quad (4.39)$$

From (4.38) and (4.39) and the estimates $L \leq \frac{A}{K}$ and $\sum_{i=1}^{R} \frac{1}{r} \leq \log R + 1$ it follows that for $2 \leq K \leq A$ and $R \geq 2$ we have

$$\sum_{0\leq i<L} \frac{1}{f'(a_{i+1}) - f'(a_i)} \int_{f'(a_i)}^{f'(a_{i+1})} T_1(\alpha, i)\,\mathrm{d}\alpha$$
$$\ll \frac{A}{K}\left(f''(A)K^3 + \frac{K}{R}\right) + \frac{A\log K(\log R+1)}{K}$$
$$\ll A\left(f''(A)K^2 + \frac{1}{R} + \frac{\log K \log R}{K}\right), \quad (4.40)$$

which concludes our treatment of the first term in (4.37). We turn to the second summand. Again we use (4.34) and obtain the estimates

$$\frac{1}{f'(a_{i+1}) - f'(a_i)} \ll \frac{1}{f''(A)K} = \frac{A}{K}\frac{1}{Af''(A)} \ll A\frac{1}{f'(2A) - f'(A)}\frac{1}{K}$$

for $0 \leq i < L$. By inserting this and the definition of $T_2(\alpha, i)$, we easily obtain

$$\sum_{0\leq i<L} \frac{1}{f'(a_{i+1}) - f'(a_i)} \int_{f'(a_i)}^{f'(a_{i+1})} T_2(\alpha, i)\,\mathrm{d}\alpha \ll A\,I(A, K). \quad (4.41)$$

To estimate the third term in (4.37), assume that $0 \leq i < L$ and $\alpha \in \left[f'(a_i), f'(a_{i+1})\right]$. We use Lemma 4.18 (setting $x = a_{i+1}$) to get

$$|a_{i+1}\alpha + f(a_i) - a_i\alpha - f(a_{i+1})| \leq c_2 f''(A)K^2,$$

therefore the two sums in the definition of $T_3(\alpha, i)$ differ by not more than $c_2 f''(A)K^2 + 1$ summands. Moreover, we have $L \leq \frac{A}{K}$. Estimating $\frac{1}{\alpha} \leq \frac{1}{f'(A)}$ we get

$$\sum_{0\leq i<L} \frac{1}{f'(a_{i+1}) - f'(a_i)} \int_{f'(a_i)}^{f'(a_{i+1})} T_3(\alpha, i)\,\mathrm{d}\alpha$$
$$\ll \frac{A}{K}\frac{1}{f'(A)}(f''(A)K^2 + 1) = A\left(\frac{f''(A)K}{f'(A)} + \frac{1}{f'(A)K}\right). \quad (4.42)$$

Finally let $0 \leq i < L$, $\alpha \in f'([a_i, a_{i+1}])$ and $f(a_i) < m \leq f(a_{i+1})$. Choose $x, y \in [a_i, a_{i+1}]$ in such a way that $\alpha = f'(x)$ and $m = f(y)$. Then by (4.34) and the monotonicity of f' we have

$$\left|\frac{1}{\alpha} - \left(f^{-1}\right)'(m)\right| = \left|\frac{1}{f'(x)} - \frac{1}{f'(y)}\right| = \left|\frac{f'(y) - f'(x)}{f'(x)f'(y)}\right|$$

$$\leq \frac{f'(a_{i+1}) - f'(a_i)}{f'(a_i)^2} \ll \frac{f''(A)K}{f'(A)^2}.$$

Moreover, the length of summation in the definition of $T_4(\alpha, i)$ can be estimated using (4.33), giving $f(a_{i+1}) - f(a_i) + 1 \ll f'(A)K + 1$. It follows that

$$\sum_{0\leq i<L} \frac{1}{f'(a_{i+1}) - f'(a_i)} \int_{f'(a_i)}^{f'(a_{i+1})} T_4(\alpha, i)\,\mathrm{d}x \ll \frac{A}{K}\left(f'(A)K + 1\right)\left(\frac{f''(A)K}{f'(A)^2}\right) \ll A\left(\frac{f''(A)K}{f'(A)} + \frac{f''(A)}{f'(A)^2}\right). \quad (4.43)$$

We still have to take care of the first and the last interval in (4.35). To do this, we take any interval $(a, b]$ such that $A \leq a \leq b \leq a + K \leq 2A$. For all $m \in \big(f(a), f(b)\big]$ we have $\left(f^{-1}\right)'(m) = \frac{1}{f'(f^{-1}(m))} \leq \frac{1}{f'(A)}$ since f' is monotonic, moreover $f(b) - f(a) + 1 \ll f'(A)K + 1 \ll f'(A)K$ by (4.33) and the relation $f'(A) \geq f'(2) > 0$, and finally $b - a + 1 \ll K$. Therefore

$$\left|\sum_{a<n\leq b} \varphi\left(\lfloor f(n)\rfloor\right) - \sum_{f(a)<m\leq f(b)} \varphi(m)\left(f^{-1}\right)'(m)\right| \ll K + f'(A)K\frac{1}{f'(A)} \ll K. \quad (4.44)$$

Combining (4.37), (4.40), (4.41), (4.42), (4.43) and (4.44) we get

$$\left|\sum_{A<n\leq 2A} \varphi\left(\lfloor f(n)\rfloor\right) - \sum_{f(A)<m\leq f(2A)} \varphi(m)\left(f^{-1}\right)'(m)\right| \ll A\left(f''(A)K^2 + \frac{1}{R} + \frac{\log K \log R}{K} + I(A, K) + \frac{f''(A)K}{f'(A)} + \frac{1}{f'(A)K} + \frac{f''(A)}{f'(A)^2} + \frac{K}{A}\right)$$

for $A, K, R \geq 2$. Since $f'(A) \geq f'(2) \gg 1$, the first term dominates the fifth and seventh terms and the third term dominates the sixth. Since $Af''(A) \gg 2f''(2) \gg 1$ by (4.29), we have $f''(A) \gg \frac{1}{A}$, and therefore the first term also dominates the last term. We choose $R = A$. Then the third term dominates the second, and the error is

$$\ll A\left(f''(A)K^2 + \frac{(\log A)^2}{K} + I(A, K)\right).$$

□

4.4.2 Proof of Theorem 4.1

We want to find an estimate for (4.8); more precisely, we want to treat the expression

$$\sum_{a<n\leq b} \varphi\left(\lfloor n\alpha + \beta\rfloor\right)$$

with the help of exponential sums. To do this, we resort to a useful approximation of the sawtooth function $x \mapsto \{x\} - \frac{1}{2}$ by trigonometric polynomials that was given by Vaaler. (See [30, Theorem A.6].)

Lemma 4.21. *Assume that H is a positive integer. There exist real numbers $a_H(h) \in [0,1]$ for $1 \le |h| \le H$ such that*

$$|\psi(t) - \psi_H(t)| \le \kappa_H(t) \tag{4.45}$$

for all real t, where

$$\psi(x) = \{x\} - \frac{1}{2},$$
$$\psi_H(t) = -\frac{1}{2\pi i} \sum_{1 \le |h| \le H} \frac{a_H(h)}{h} \, \mathrm{e}(ht)$$

and

$$\kappa_H(t) = \frac{1}{2(H+1)} \sum_{0 \le |h| \le H} \left(1 - \frac{|h|}{H+1}\right) \mathrm{e}(ht).$$

Note that $\kappa_H(t)$ is a nonnegative real number since for all H we have

$$\sum_{0 \le |h| < H} (H - |h|) e(hx) = \left| \sum_{0 \le h < H} \mathrm{e}\,(hx) \right|^2.$$

Let α and β be real numbers and suppose that $\alpha \ge 1$. An elementary argument shows that for all integers m we have

$$\left\lfloor -\frac{m-\beta}{\alpha} \right\rfloor - \left\lfloor -\frac{m+1-\beta}{\alpha} \right\rfloor = \left\{ \begin{array}{ll} 1 & \text{if } m = \lfloor n\alpha + \beta \rfloor \text{ for some integer } n \\ 0 & \text{otherwise.} \end{array} \right\}. \tag{4.46}$$

With the help of this characterization of the elements of a Beatty sequence we prove the following statement, which allows us to deduce Theorem 4.1 from Proposition 4.2.

Proposition 4.22. *Let $\varphi : \mathbb{N} \to \mathbb{C}$ be a function bounded by 1. For all real $\alpha \ge 1$, $\beta \ge 0$, $K \ge 0$ and $H \ge 1$ we have*

$$\left| \sum_{0 < n \le K} \varphi\left(\lfloor n\alpha + \beta \rfloor\right) - \frac{1}{\alpha} \sum_{\beta < m \le \beta + K\alpha} \varphi(m) \right| \le \sum_{1 \le |h| \le H} \min\left\{ \frac{1}{\alpha}, \frac{1}{|h|} \right\} \left| \sum_{\beta < m \le \beta + K\alpha} \varphi(m) \, \mathrm{e}\left(-m\frac{h}{\alpha}\right) \right| + \frac{1}{H} \sum_{0 \le |h| \le H} \left| \sum_{\beta < m \le \beta + K\alpha} \mathrm{e}\left(-m\frac{h}{\alpha}\right) \right| + O(1).$$

The implied constant is an absolute one.

Proof. We write $\psi(x) = \{x\} - \frac{1}{2} = x - \lfloor x \rfloor - \frac{1}{2}$. Since $\alpha \geq 1$, the function $n \mapsto \lfloor n\alpha + \beta \rfloor$ is injective. Using this fact and (4.46), we see that

$$\begin{aligned}
&\sum_{0<n\leq K} \varphi\left(\lfloor n\alpha+\beta\rfloor\right) \\
&\qquad = \sum_{m\in\mathbb{Z}} \varphi(m) \cdot \left\{ \begin{array}{ll} 1 & m = \lfloor n\alpha+\beta\rfloor \text{ for some } 0<n\leq K \\ 0 & \text{otherwise} \end{array} \right\} \\
&\qquad = \sum_{\lfloor\beta\rfloor<m\leq\lfloor\beta+K\alpha\rfloor} \varphi(m) \cdot \left\{ \begin{array}{ll} 1 & m = \lfloor n\alpha+\beta\rfloor \text{ for some } n \\ 0 & \text{otherwise} \end{array} \right\} \\
&\qquad = \sum_{\lfloor\beta\rfloor<m\leq\lfloor\beta+K\alpha\rfloor} \varphi(m) \left(\left\lfloor -\frac{m-\beta}{\alpha} \right\rfloor - \left\lfloor -\frac{m+1-\beta}{\alpha} \right\rfloor \right) \\
&\qquad\qquad = \frac{1}{\alpha} \sum_{\beta<m\leq\beta+K\alpha} \varphi(m) \\
&\qquad\qquad\quad + \sum_{\beta<m\leq\beta+K\alpha} \varphi(m) \left(\psi\left(-\frac{m+1-\beta}{\alpha} \right) - \psi\left(-\frac{m-\beta}{\alpha} \right) \right) + O(1).
\end{aligned}$$

It remains to treat the second sum. For brevity, write

$$L = \{m \in \mathbb{Z} : \beta < m \leq \beta + K\alpha\}.$$

Let $H \geq 1$ be an integer. For each m we replace ψ by ψ_H with the help of (4.45) to get

$$\begin{aligned}
&\left| \sum_{m\in L} \varphi(m) \left(\psi\left(-\frac{m+1-\beta}{\alpha} - \gamma \right) - \psi\left(-\frac{m-\beta}{\alpha} - \gamma \right) \right) \right. \\
&\qquad \left. - \frac{-1}{2\pi i} \sum_{m\in L} \varphi(m) \sum_{1\leq|h|\leq H} \frac{a_H(h)}{h} \left(\mathrm{e}\left(-h\frac{m+1-\beta}{\alpha} \right) - \mathrm{e}\left(-h\frac{m-\beta}{\alpha} \right) \right) \right| \\
&\qquad \leq \frac{1}{2H+2} \sum_{m\in L} \sum_{|h|\leq H} \left(1 - \frac{|h|}{H+1} \right) \left(\mathrm{e}\left(-h\frac{m+1-\beta}{\alpha} \right) + \mathrm{e}\left(-h\frac{m-\beta}{\alpha} \right) \right) \\
&\qquad\qquad \leq \frac{1}{H+1} \sum_{0\leq|h|\leq H} \left| \sum_{m\in L} \mathrm{e}\left(-h\frac{m}{\alpha} \right) \right|.
\end{aligned}$$

Finally we use the inequalities $|a_H(h)| \leq 1$ and $|\mathrm{e}(x) - 1| \leq \min\{2, 2\pi x\}$ to calculate:

$$\begin{aligned}
&\left| \frac{1}{2\pi i} \sum_{m\in L} \varphi(m) \sum_{1\leq|h|\leq H} \frac{a_H(h)}{h} \left(\mathrm{e}\left(-h\frac{m+1-\beta}{\alpha} \right) - \mathrm{e}\left(-h\frac{m-\beta}{\alpha} \right) \right) \right| \\
&\qquad = \left| \frac{1}{2\pi} \sum_{1\leq|h|\leq H} \frac{a_H(h)}{h} \mathrm{e}\left(-\frac{\beta}{\alpha} \right) \left(\mathrm{e}\left(-\frac{h}{\alpha} \right) - 1 \right) \sum_{m\in L} \varphi(m) \left(-h\frac{m}{\alpha} \right) \right|
\end{aligned}$$

$$\leq \sum_{1\leq|h|\leq H} \min\left\{\frac{1}{\alpha}, \frac{1}{|h|}\right\} \left|\sum_{m\in L} \varphi(m)\left(-h\frac{m}{\alpha}\right)\right|.$$

If $H \geq 1$ is a real number, we apply these calculations to $\lfloor H \rfloor$. Note that in this process the summations over h remain unchanged and that $1/(\lfloor H \rfloor + 1) \leq 1/H$, therefore the assertion follows. □

We will use the following standard lemma to extend the range of a summation in exchange for a controllable factor.

Lemma 4.23. *Let $x \leq y \leq z$ be real numbers and $a_n \in \mathbb{C}$ for $x < n \leq z$. Then*

$$\left|\sum_{x<n\leq y} a_n\right| \leq \int_0^1 \min\left\{y - x + 1, \|\xi\|^{-1}\right\} \left|\sum_{x<n\leq z} a_n \,\mathrm{e}\,(n\xi)\right| \mathrm{d}\xi.$$

Proof. Since $\int_0^1 \mathrm{e}\,(k\xi)\ \mathrm{d}\xi = \delta_{k,0}$ for $k \in \mathbb{Z}$ it follows that

$$\sum_{x<n\leq y} a_n = \sum_{x<n\leq z} a_n \sum_{x<m\leq y} \delta_{n-m,0} = \int_0^1 \sum_{x<m\leq y} \mathrm{e}\,(-m\xi) \sum_{x<n\leq z} a_n \,\mathrm{e}\,(n\xi)\ \mathrm{d}\xi,$$

from which the statement follows. □

Finally, to obtain the correct error term in the theorem, we will use the following lower bound on the L^1-norm of an exponential sum.

Lemma 4.24. *Let $a < b$ be real numbers and x_m a complex number for $a < m \leq b$. Then*

$$\int_0^1 \left|\sum_{a<m\leq b} x_m \,\mathrm{e}\,(m\vartheta)\right| \mathrm{d}\vartheta \geq \max_{a<m\leq b} |x_m|\,.$$

Proof. For $a < n \leq b$ we have

$$\int_0^1 \left|\sum_{a<m\leq b} x_m \,\mathrm{e}\,(m\vartheta)\right| \mathrm{d}\vartheta = \int_0^1 \left|\sum_{a<m\leq b} x_m \,\mathrm{e}\,((m-n)\vartheta)\right| \mathrm{d}\vartheta$$
$$\geq \left|\sum_{a<m\leq b} x_m \int_0^1 \mathrm{e}\,((m-n)\vartheta)\ \mathrm{d}\vartheta\right| = x_n.$$

□

Proof of Theorem 4.1. Note first that by (4.32) we have $f'(x) \to \infty$, therefore there exists $A_0 \geq 2$ such that $f'(A) \geq 1$ for $A \geq A_0$. Let $z > 0$. By an argument similar to that at the beginning of the proof of Proposition 4.2 we may restrict ourselves to the case that $z \leq A\, f'(A)$. Also, we may assume that there exists an m in the range $f(A) < m \leq f(2A) + z$ such that $|\varphi(m)| = 1$, since the general case follows from this one by rescaling both sides of (4.5). To see this, we note that $A \geq 2$ is an integer and $f'(x) \geq 1$ for all $x \geq A$ and

therefore the relation (4.5) only depends on integers m in the range $f(A) < m \le f(2A) + z$. By Lemma 4.24, this restriction implies

$$\int_0^1 \sup_{f(A)<x\le f(2A)} \left| \sum_{x<m\le x+z} \varphi(m)\,\mathrm{e}\,(m\vartheta) \right| \mathrm{d}\vartheta$$
$$\ge \sup_{f(A)<x\le f(2A)} \int_0^1 \left| \sum_{x<m\le x+z} \varphi(m)\,\mathrm{e}\,(m\vartheta) \right| \mathrm{d}\vartheta \ge \sup_{f(A)<x\le f(2A)} \sup_{x<m\le x+z} |\varphi(m)|$$
$$= \sup_{f(A)<m\le f(2A)+z} |\varphi(m)| \ge 1. \quad (4.47)$$

If $z < \max\{2, f'(2A)\}$, this lower bound implies $f'(A)(\log A)^3 J(A,z) \gg 1$ and since by Lemma 4.15 the left hand side of (4.5) is bounded, this proves the assertion in this case. For the remaining part of the proof we assume therefore that $\max\{2, f'(2A)\} \le z \le A\,f'(A)$. Moreover, we assume throughout that $1 \le K \le A$ and that $H \ge 2$. We want to apply Proposition 4.2 and therefore we have to find an estimate for $I(A,K)$. We apply Proposition 4.22 to the expression in the absolute value in equation (4.8), which is possible since $\alpha \ge f'(A) \ge 1$ for all α in question, and obtain the estimate

$$I(A,K) \ll \frac{1}{f'(2A)-f'(A)} \frac{1}{K} \int_{f'(A)}^{f'(2A)} \left(\sum_{1\le|h|\le H} \min\left\{ \frac{1}{\alpha}, \frac{1}{|h|} \right\} S_1(\alpha,h) \right.$$
$$\left. + \frac{1}{H} S_2(\alpha,0) + \frac{1}{H} \sum_{1\le|h|\le H} S_2(\alpha,h) + O(1) \right) \mathrm{d}\alpha, \quad (4.48)$$

where

$$S_1(\alpha,h) = \sup_{f(A)<x\le f(2A)} \left| \sum_{x<m\le x+K\alpha} \varphi(m)\,\mathrm{e}\left(-m\frac{h}{\alpha}\right) \right|$$

and

$$S_2(\alpha,h) = \sup_{f(A)<x\le f(2A)} \left| \sum_{x<m\le x+K\alpha} \mathrm{e}\left(-m\frac{h}{\alpha}\right) \right|.$$

The four summands in (4.48) are arranged according to their importance. We estimate them in the order of increasing importance, the treatment of the fourth term being trivial:

$$\int_{f'(A)}^{f'(2A)} O(1)\,\mathrm{d}\alpha \ll f'(2A) - f'(A). \quad (4.49)$$

To estimate the third term, it is sufficient to consider the sum over $1 \le h \le H$, since $S_2(\alpha,-h) = S_2(\alpha,h)$. We interchange the integration and the summation and substitute $\vartheta = -\frac{h}{\alpha}$ to obtain

$$\int_{f'(A)}^{f'(2A)} \frac{1}{H} \sum_{1\le h\le H} S_2(\alpha,h)\,\mathrm{d}\alpha$$

$$\ll \frac{1}{H} \sum_{1 \le h \le H} h \int_{-\frac{h}{f'(A)}}^{-\frac{h}{f'(2A)}} \frac{1}{\vartheta^2} \min\left\{ f'(2A)K + 1, \|\vartheta\|^{-1} \right\} \mathrm{d}\vartheta.$$

We note some simple estimates before applying Lemma 4.17. We have $0 < f'(1) \le f'(2A) \ll A^\delta$ for some $\delta \ge 0$ since f' is monotone and by (4.32), and therefore $f'(2A)K + 1 \ll A^{\delta+1}$. By (4.31) we have $0 < -\frac{1}{\vartheta} \le \frac{f'(2A)}{h} \ll \frac{f'(A)}{h}$ for all ϑ under consideration. Moreover, the length of the integration range is $\frac{h}{f'(A)} - \frac{h}{f'(2A)} \le \frac{h}{f'(A)}$ and finally from (4.30) and (4.34) it follows that $f'(A) \ll (f'(2A) - f'(A)) \log A$. Hence Lemma 4.17 gives

$$\begin{aligned}
&\int_{f'(A)}^{f'(2A)} \frac{1}{H} \sum_{1 \le h \le H} S_2(\alpha, h) \, \mathrm{d}\alpha \\
&\qquad \ll f'(A) \frac{1}{H} \sum_{1 \le h \le H} \frac{f'(A)}{h} \left(\frac{h}{f'(A)} + 1 \right) \left(1 + \log A^{\delta+1} \right) \\
&\qquad \ll f'(A) \left(1 + \frac{f'(A) \log H}{H} \right) \log A \\
&\qquad \ll \left(f'(2A) - f'(A) \right) (\log A)^2 \left(1 + \frac{f'(A) \log H}{H} \right). \quad (4.50)
\end{aligned}$$

The contribution of the second term in (4.48) is easily determined: the sum occurring in the definition of S_2 comprises not more than $f'(2A)K + 1 \ll f'(A)K$ summands, therefore

$$\int_{f'(A)}^{f'(2A)} \frac{1}{H} S_2(\alpha, 0) \, \mathrm{d}\alpha \ll (f'(2A) - f'(A))K \frac{f'(A)}{H}. \quad (4.51)$$

Now we turn to the the treatment of the main term in (4.48). We concentrate on the case that $h > 0$. We exchange the integral and the sum and apply the substitution $-\frac{h}{\alpha} = \vartheta$. The factor $\min\left\{\frac{1}{\alpha}, \frac{1}{h}\right\}$ then transforms into $\min\left\{-\frac{1}{\vartheta}, \frac{1}{\vartheta^2}\right\}$, which is $\ll \frac{f'(A)}{h} \min\left\{1, \frac{f'(A)}{h}\right\}$ by (4.31). We obtain

$$\begin{aligned}
&\int_{f'(A)}^{f'(2A)} \sum_{1 \le h \le H} \min\left\{ \frac{1}{\alpha}, \frac{1}{h} \right\} S_1(\alpha, h) \, \mathrm{d}\alpha \\
&\qquad \ll f'(A) \sum_{1 \le h \le H} \frac{1}{h} \min\left\{ 1, \frac{f'(A)}{h} \right\} \int_{-\frac{h}{f'(A)}}^{-\frac{h}{f'(2A)}} S_1\left(-\frac{h}{\vartheta}, h \right) \mathrm{d}\vartheta
\end{aligned}$$

and to estimate the integral we use Lemma 4.23:

$$\begin{aligned}
\int_{-\frac{h}{f'(A)}}^{-\frac{h}{f'(2A)}} S_1\left(-\frac{h}{\vartheta}, h \right) \mathrm{d}\vartheta &\ll \int_0^1 \min\left\{ f'(2A)K + 1, \|\xi\|^{-1} \right\} \\
&\quad \times \int_{\xi - \frac{h}{f'(A)}}^{\xi - \frac{h}{f'(2A)}} \sup_{f(A) < x \le f(2A)} \left| \sum_{x < m \le x + f'(2A)K} \varphi(m) \, \mathrm{e}\left(m\vartheta \right) \right| \mathrm{d}\vartheta \, \mathrm{d}\xi.
\end{aligned}$$

The length of integration of the inner integral is bounded trivially by $\frac{h}{f'(A)}$ and the integrand is 1-periodic, so that we may replace this integral, using the definition (4.6) of J, by the upper bound

$$\left(\frac{h}{f'(A)}+1\right) f'(2A)K\, J\big(A, f'(2A)K\big),$$

which is independent of ξ. We use the estimate $f'(2A)K+1 \ll A^{\delta+1}$ which we mentioned before and Lemma 4.17 to obtain

$$\int_0^1 \min\left\{f'(2A)K+1, \|\xi\|^{-1}\right\} \mathrm{d}\xi \ll \log A.$$

Splitting the summation over h at $f'(A)$ we get

$$\sum_{1\le h\le H} \frac{1}{h}\min\left\{1, \frac{f'(A)}{h}\right\}\left(\frac{h}{f'(A)}+1\right)$$
$$\ll \sum_{1\le h\le f'(A)} \frac{1}{h} + \sum_{f'(A)<h\le H} \frac{1}{h}\frac{f'(A)}{h}\frac{h}{f'(A)} \ll \sum_{1\le h\le H}\frac{1}{h} \ll \log H.$$

Collecting the terms and using the estimate $f'(2A) \ll \big(f'(2A)-f'(A)\big)\log A$, which follows from Lemma 4.16, we get

$$\int_{f'(A)}^{f'(2A)} \sum_{1\le h\le H} \min\left\{\frac{1}{\alpha}, \frac{1}{h}\right\} S_1(\alpha, h)\,\mathrm{d}\alpha$$
$$\ll f'(A)\big(f'(2A)-f'(A)\big)K(\log A)^2 \log H\, J\big(A, f'(2A)K\big). \quad (4.52)$$

By analogous reasoning the sum over $-H \le h \le -1$ can be estimated by the same expression. We choose

$$H = z \quad \text{and} \quad K = \frac{z}{f'(2A)}.$$

By the restrictions $\max\{2, f'(2A)\} \le z \le A\, f'(A)$ it easily follows that $1 \le K \le A$ and that $H \ge 2$, therefore this is an admissible choice. Note also that $\log H \ll \log A$ by (4.32). We combine (4.48), (4.49), (4.50), (4.51) and (4.52) to get the estimate

$$I\left(A, \frac{z}{f'(2A)}\right) \ll \frac{f'(A)(\log A)^3}{z} + f'(A)(\log A)^3 J(A, z).$$

Applying Proposition 4.2 we see that the left hand side of (4.5) is bounded by a constant times

$$\frac{f''(A)}{f'(A)^2}z^2 + \frac{f'(A)(\log A)^3}{z} + f'(A)(\log A)^3 J(A, z).$$

By (4.47) the second term in this expression is dominated by the third, which completes the proof. □

Chapter 5

The Zeckendorf sum-of-digits function

In the previous chapters we have encountered the Zeckendorf sum-of-digits Z twice: in section 1.2.4 we have proved that for $\beta \in \mathbb{R} \setminus \mathbb{Z}$ the arithmetic function $n \mapsto \mathrm{e}(\beta Z(n))$ is pseudorandom (in the sense of Bertrandias). Moreover, in section 4.3.3 we have studied the distribution of the values of $Z(m)$ in residue classes, where m is of the form $\lfloor n^c \rfloor$.

In the present chapter we investigate the relation of the Zeckendorf sum-of-digits function to ordinary sum-of-digits functions, showing that their values distribute independently in residue classes. A non-quantitative version of this result was proved by Coquet, Rhin and Toffin in 1981 [13, Théorème 3]. Our result is new insofar as it provides an explicit error term. This result is to be compared to the results of Bésineau [7] and Kim [33], which treat the case of ordinary q-ary representations in different (coprime) bases. Our proof uses (again) van der Corput's inequality followed by a discrete Fourier transform, which has proven to be a successful combination in [41, 42], see also section 1.3 and section 4.3.2. However, in order to treat the function Z, we have to use modified Fourier coefficients and a uniform estimate for these terms. Such an estimate was derived in Lemma 1.31.

5.1 Introduction and main results

The problem of studying the joint distribution of sum-of-digits functions in different (coprime) bases was rised by Gelfond [29]. As we discussed before (see section 1.3 and section 4.1), this problem has been solved completely. In the article [13] a variant of this problem is attacked and the following theorem is proved.

Theorem 5.1 (Coquet, Rhin, Toffin)**.** *Let α be irrational and $q \geq 2$ and let $\sigma_\alpha(n)$ be the α-sum-of-digits of n. The sequence $n \mapsto x s_q(n) + y\sigma_\alpha(n)$ is uniformly distributed modulo 1 if and only if at least one of x and y is irrational.*

We do not define the α-sum-of-digits here, we only note that it is defined via the Ostrowski expansion of an integer n with respect to α, of which the Zeckendorf expansion (1.21) is a special case, taking $\alpha = \varphi$. We recall that $Z(n) = \sum_{i \geq 2} \varepsilon_i$, where $n = \sum_{i \geq 2} \varepsilon_i F_i$ and the coefficients $\varepsilon_i \in \{0, 1\}$ satisfy the restriction $\varepsilon_i = 1 \Rightarrow \varepsilon_{i+1} = 0$.

The above theorem is a statement on uniform distribution, in particular it is a non-quantitative statement. We (partly) improve it by achieving an error term in the case of

the Zeckendorf sum-of-digits. The purpose of this short chapter is to prove the following theorem.

Theorem 5.2. *Let $q \geq 2$ be an integer and ϑ, β be real numbers such that $\beta \notin \mathbb{Z}$. Then*

$$\sum_{n<N} \mathrm{e}\left(\vartheta s_q(n) + \beta Z(n)\right) = O\left(N^{1-\eta}\right)$$

for some $\eta > 0$.

As usual, we can derive a statement on joint distribution in residue classes from this theorem.

Corollary 5.3. *Let $q \geq 2$ be an integer. Assume that $m_1 \geq 1$, $m_2 \geq 2$ are integers and that $(q-1, m_1) = 1$. There exists an $\eta > 0$ such that for all $a_1, a_2 \in \mathbb{Z}$ we have*

$$\frac{1}{x}\left|\{n < x : s_q(n) \equiv a_1 \bmod m_1, Z(n) \equiv a_2 \bmod m_2\}\right| = \frac{1}{m_1 m_2} + O(x^{-\eta}).$$

We note that in particular the Sequence Z is uniformly distributed in $\mathbb{Z}$ (that is, uniformly distributed in each residue class, since there is no restriction on the modulus $m_2 \geq 2$. This behaviour is different from s_q (where $q \geq 3$), which is not uniformly distributed in $(q-1)\mathbb{Z}$ by the congruence $s_q(n) \equiv n \bmod q-1$.

Throughout this chapter, we denote the golden ratio by φ, that is, $\varphi = \frac{1}{2}(\sqrt{5}+1)$. The sequence of Fibonacci numbers $(F_i)_{i\geq 0} = (0, 1, 1, 2, \ldots)$ satisfies the well-known property that

$$F_i = \frac{\varphi^i + (-1)^{i+1}\varphi^{-i}}{\sqrt{5}} \tag{5.1}$$

for $i \geq 0$. The main new tool that we use in the proof of Theorem 5.2 is the following representation of the expression $\mathrm{e}\left(\vartheta Z_k(n)\right)$, where $Z_k(n) = \sum_{2\leq i<k} \varepsilon_i(n)$ for $k \geq 2$. This proposition is the analogue of the much simpler discrete Fourier transform for the q^λ-periodic function $\mathrm{e}\left(\vartheta s_{q,\lambda}\right)$, see Definition 1.10 and the equations following it.

Proposition 5.4. *Assume that $k \geq 2$ and $H \geq 1$. Let*

$$M_k^{(1)}(h, \vartheta) = \sum_{u<F_{k-1}} \mathrm{e}\left(\vartheta Z(u) - (-1)^k h u \varphi\right)$$
$$M_k^{(2)}(h, \vartheta) = \sum_{F_{k-1}\leq u<F_k} \mathrm{e}\left(\vartheta Z(u) - (-1)^k h u \varphi\right).$$

There exist complex numbers $b_H^{(1)}(h), b_H^{(2)}(h), c_H^{(1)}(h)$ and $c_H^{(2)}(h)$ for $|h| \leq H$ such that

$$\begin{aligned}
b^{(1)}(0) &= \varphi^{-k+2},\\
\left|b^{(1)}(h)\right| &\leq \min\left\{\frac{2}{|h|}, \varphi^{-k+2}\right\},\\
b^{(2)}(0) &= \varphi^{-k+1},\\
\left|b^{(2)}(h)\right| &\leq \min\left\{\frac{2}{|h|}, \varphi^{-k+1}\right\},\\
\left|c^{(1)}(h)\right| &\leq 2,\\
\left|c^{(2)}(h)\right| &\leq 2
\end{aligned}$$

and

$$\begin{aligned}\mathrm{e}\left(\vartheta Z_k(n)\right) = &\sum_{|h|\le H} \mathrm{e}\left((-1)^k hn\varphi\right) b_H^{(1)}(h) M_k^{(1)}(h,\vartheta)\\ &+ \sum_{|h|\le H} \mathrm{e}\left((-1)^k hn\varphi\right) b_H^{(2)}(h) M_k^{(2)}(h,\vartheta)\\ &+ O\left(\frac{1}{H}\sum_{|h|\le H} c_H^{(1)}(h)\,\mathrm{e}\left((-1)^k hn\varphi\right) \sum_{u<F_{k-1}} \mathrm{e}\left(-(-1)^k hu\varphi\right)\right)\\ &+ O\left(\frac{1}{H}\sum_{|h|\le H} c_H^{(2)}(h)\,\mathrm{e}\left((-1)^k hn\varphi\right) \sum_{F_{k-1}\le u<F_k} \mathrm{e}\left(-(-1)^k hu\varphi\right)\right)\end{aligned} \tag{5.2}$$

for all nonnegative integers n and real numbers ϑ, where the expressions in parentheses are nonnegative real numbers and the implied constants are absolute.

The idea of the proof of this proposition lies in the fact that the integers n such that the Zeckendorf representation of n starts with a given sequence $(a_2, \dots, a_{k-1})$ in $\{0,1\}$ can be characterized by the relation $\{n\varphi\} \in I$, where I is a certain subinterval of $[0,1)$. There are intervals I of two different lengths, corresponding to the value of a_{k-1}, which explains the presence of two main terms in the proposition. Trigonometric approximation of the indicator function of $I+\mathbb{Z}$ via Lemma 4.21 will allow us to obtain the statement after some elementary manipulation.

5.2 Proofs

5.2.1 Auxiliary lemmas

For any $k \ge 2$, we will use the following function which "cuts off" the digits with indices $\ge k$:

$$v(n,k) = \sum_{2\le i<k} \varepsilon_i(n) F_i.$$

Note that this is the counterpart to the function $n \mapsto n \bmod q^k$ in the case of the q-ary representation of integers. We define the sum-of-digits function of the digits up to k:

$$Z_k(n) = Z(v(n,k)) = \sum_{2\le i<k} \varepsilon_i(n).$$

We want to detect in an analytical way whether $v(n,k)$ has a given value. In order to do so, we first prove the following lemma.

Lemma 5.5. *Let n be a nonnegative integer, $k \ge 2$ and $v = v(n,k) = \sum_{i=2}^{k-1} \varepsilon_i(n) F_i$. We define*

$$A_k^{(1)} = \left[-\frac{1}{\varphi^{k-1}}, \frac{1}{\varphi^k}\right) \quad \text{and} \quad A_k^{(2)} = \left[-\frac{1}{\varphi^{k+1}}, \frac{1}{\varphi^k}\right).$$

Let $p_k(n) = (-1)^k n\varphi$. Moreover, we set

$$R_k(u) = p_k(u) + \begin{cases} A_k^{(1)}, & 0 \le u < F_{k-1} \\ A_k^{(2)}, & F_{k-1} \le u < F_k. \end{cases} \tag{5.3}$$

Then

$$p_k(n) \in R_k(v(n,k)) + \mathbb{Z}.$$

Proof. We use (5.1) and calculate:

$$\begin{aligned} n\varphi &= v(n,k)\varphi + \sum_{i\ge k} \varepsilon_i \varphi \frac{\varphi^i - (-\varphi)^{-i}}{\sqrt{5}} \\ &= v(n,k)\varphi + \sum_{i\ge k} \varepsilon_i \frac{\varphi^{i+1} - (-\varphi)^{-(i+1)}}{\sqrt{5}} + \sum_{i\ge k} \varepsilon_i \frac{-(-\varphi)^{-i}\varphi + (-\varphi)^{-(i+1)}}{\sqrt{5}}. \end{aligned}$$

We note that the second term is a sum of Fibonacci numbers and as such is an integer. Moreover, we have the identity $(1+\varphi^{-2})/\sqrt{5} = 1/\varphi$. Therefore

$$n\varphi \equiv v(n,k)\varphi + \sum_{i\ge k} \varepsilon_i \frac{(-\varphi)^{-i+1}\,(1+(-\varphi)^{-2})}{\sqrt{5}} \equiv v(n,k)\varphi - \sum_{i\ge k} \varepsilon_i \frac{1}{(-\varphi)^i} \bmod 1.$$

We write

$$s(n,k) = -\sum_{i\ge k} \varepsilon_i \frac{1}{(-\varphi)^i}.$$

We derive bounds for $s(n,k)$. Assume first that k is even and that $\varepsilon_{k-1} = 0$. Then we certainly have

$$\sum_{i\ge k} \varepsilon_i \frac{1}{(-\varphi)^i} < \sum_{\ell\ge 0} \frac{1}{(-\varphi)^{k+2\ell}}$$

since in the sum on the right hand side we take all the positive summands. It follows that

$$s(n,k) > -\frac{1}{\varphi^k} \cdot \frac{1}{1-\varphi^{-2}} = -\frac{1}{\varphi^{k-1}}.$$

Similarly, we have

$$s(n,k) < \frac{1}{\varphi^k}.$$

In the case that $\varepsilon_{k-1} = 0$ we cannot choose $\varepsilon_k = 1$. Therefore we obtain in this case

$$-\frac{1}{\varphi^{k+1}} < s(n,k) < \frac{1}{\varphi^k}.$$

In an analogous way we treat the case that k is odd. In this case we obtain

$$n(-\varphi) \equiv v(n,k)(-\varphi) + \sum_{i\ge k} \varepsilon_i \frac{1}{(-\varphi)^i} \bmod 1.$$

We define

$$s^-(n,k) = -s(n,k) = \sum_{i \geq k} \varepsilon_i \frac{1}{(-\varphi)^i}.$$

We obtain by a similar computation as above

$$-\frac{1}{\varphi^{k-1}} < s^-(n,k) < \frac{1}{\varphi^k}$$

if $\varepsilon_{k-1} = 0$ and

$$-\frac{1}{\varphi^{k+1}} < s^-(n,k) < \frac{1}{\varphi^k}$$

if $\varepsilon_{k-1} = 1$.

□

In order to characterize the integers n with a given initial segment $v(n,k)$ in the Zeckendorf representation, we have to show that the intervals $R_k(u)$ are disjoint modulo one.

Lemma 5.6. *The sets*

$$R_k(u) + \mathbb{Z},$$

where $0 \leq u < F_k$*, form a partition of* $\mathbb{R}$*.*

Proof. We set

$$\tilde{R}_k(u) = R_k(u) \bmod 1.$$

We first show that

$$\bigcup_{u<F_k} \tilde{R}_k(u) = [0,1)$$

by contradiction. Assume that $x \in [0,1)$ is such that $x \notin \tilde{R}_k(u)$ for all u. Since each $\tilde{R}_k(u)$ is the union of at most two intervals of the form $[a,b)$, there is an $\varepsilon > 0$ such that $[x, x+\varepsilon] \cap \bigcup \tilde{R}_k(u) = \emptyset$. Since the values $\{p_k(u)\}$ are dense in the unit interval, there exists an n such that $\{p_k(u)\} \in [x, x+\varepsilon]$. Moreover, we have $\{p_k(n)\} \in \tilde{R}_k(u)$ for some u by Lemma 5.5, which is a contradiction.

We show disjointness of the sets $\tilde{R}_k(u)$. Assume that $x \in \tilde{R}_k(v_1) \cap \tilde{R}_k(v_2)$, where $v_1 \neq v_2$. Then $\lambda(\tilde{R}_k(v_1) \cap \tilde{R}_k(v_2)) \geq \varepsilon$ for some $\varepsilon > 0$. By the identity

$$\frac{F_{k-1}}{\varphi^{k-2}} + \frac{F_{k-2}}{\varphi^{k-1}} = 1$$

the sum of the measures of $\tilde{R}_k(u)$ equals 1. We calculate:

$$\begin{aligned} 1 = \lambda\left(\bigcup_{u<F_k} \tilde{R}_k(u)\right) &= \lambda\left(\left(\tilde{R}_k(v_1) \setminus (\tilde{R}_k(v_1) \cap \tilde{R}_k(v_2))\right) \cup \bigcup_{u \neq v_1} \tilde{R}_k(u)\right) \\ &\leq \sum_{u<F_k} \lambda(\tilde{R}_k(v)) - \varepsilon = 1 - \varepsilon < 1, \end{aligned}$$

a contradiction. □

Combining Lemma 5.5 and Lemma 5.6 we obtain the following statement.

Proposition 5.7. *Let $k \geq 2, 0 \leq u < F_k$ and $n \geq 0$. Then we have*

$$v(n,k) = u$$

if and only if

$$(-1)^k n\varphi \in R_k(u) + \mathbb{Z}.$$

In other words, the first $k-2$ digits in the Zeckendorf expansion of an integer are coded by $n \mapsto n\varphi$ (resp. $n \mapsto n(-\varphi)$) together with the partition defined by the sets $R_k(u) + \mathbb{Z}$. This characterization in combination with trigonometric approximation of the indicator functions of the sets $R_k(u) + \mathbb{Z}$ will allow us to replace the combinatorial condition that the first digits of the expansion are fixed, $(\varepsilon_2(n), \ldots, \varepsilon_{k-1}(n)) = (a_2, \ldots, a_{k-1})$, by an analytical statement.

5.2.2 Proof of Proposition 5.4

We set

$$g_u = \mathbf{1}_{R_k(u)+\mathbb{Z}} \text{ for } 0 \leq u < F_k. \tag{5.4}$$

By Proposition 5.7 we have for all u, $0 \leq u < F_k$ and all $n \geq 0$

$$g_u(p_k(n)) = \begin{cases} 1 & \text{if } u = v(n,k) \\ 0 & \text{otherwise.} \end{cases} \tag{5.5}$$

Therefore we can calculate:

$$\begin{aligned} e\left(\vartheta Z_k(n)\right) &= e\left(\vartheta Z(v(n,k))\right) \\ &= \sum_{u<F_k} e\left(\vartheta Z(u)\right) \left\{ \begin{array}{ll} 1 & \text{if } u = v(n,k) \\ 0 & \text{otherwise} \end{array} \right\} \\ &= \sum_{u<f_k} e\left(\vartheta Z(u)\right) g_u(p_k(n)) \\ &= \sum_{u<F_{k-1}} e\left(\vartheta Z(u)\right) g_u(p_k(n)) + \sum_{F_{k-1}\leq u<F_k} e\left(\vartheta Z(u)\right) g_u(p_k(n)). \end{aligned} \tag{5.6}$$

Let $0 \leq a \leq b \leq 1$. We have $\{x\} \in [a,b)$ if and only if $1 = \lfloor x-a \rfloor - \lfloor x-b \rfloor$, therefore, writing $\psi(x) = \{x\} - 1/2$,

$$b - a + \psi(x-b) - \psi(x-a) = \begin{cases} 1 & \text{if } \{x\} \in [a,b) \\ 0 & \text{otherwise.} \end{cases}$$

We apply Lemma 4.21 in order to obtain exponential sums: we have

$$\begin{aligned} \mathbf{1}_{[a,b)+\mathbb{Z}}(x) &= b - a + \psi_H(x-b) - \psi_H(x-a) + O\big(\kappa_H(x-b) + \kappa_H(x-a)\big) \\ &= b - a - \frac{1}{2\pi i} \sum_{1\leq|h|\leq H} \frac{a_H(h)}{h} \left(1 - \mathrm{e}\left(h(b-a)\right)\right) \mathrm{e}\left(h(x-b)\right) \\ &\quad + O\big(\kappa_H(x-b) + \kappa_H(x-a)\big) \\ &= \sum_{|h|\leq H} a'_H(h)\, \mathrm{e}\left(h(x-b)\right) + O\big(\kappa_H(x-b) + \kappa_H(x-a)\big) \end{aligned} \tag{5.7}$$

for some $a'_H(h)$ such that $a'_H(0) = b-a$ and $|a'_H(h)| \le \min\{b-a, 1/|h|\}$ for $h \neq 0$. We treat the first sum in (5.6), the second being analogous. We use (5.7) and (5.3) to obtain

$$\begin{aligned}\sum_{u<F_{k-1}} e(\vartheta Z(u))\, g_u(p_k(n)) &= \sum_{u<F_{k-1}} \mathrm{e}(\vartheta Z(u)) \sum_{|h|\le H} a'_H(h)\, \mathrm{e}\left(h(p_k(n)-p_k(u)-\varphi^{-k})\right)\\ &\quad + O\left(\sum_{u<F_{k-1}} \mathrm{e}(\vartheta Z(u))\, \kappa_H(p_k(n)-p_k(u)-\varphi^{-k})\right)\\ &\quad + O\left(\sum_{u<F_{k-1}} \mathrm{e}(\vartheta Z(u))\, \kappa_H(p_k(n)-p_k(u)+\varphi^{-k+1})\right)\\ &= M + O(E_1) + O(E_2).\end{aligned}$$

For the main term we have

$$M = \sum_{|h|\le H} \mathrm{e}\left((-1)^k hn\varphi\right) b_H(h) \sum_{u<F_{k-1}} \mathrm{e}\left(\vartheta Z(u) - (-1)^k hu\varphi\right),$$

where

$$\begin{aligned} b_H(0) &= \varphi^{-k+2}\\ b_H(h) &= a'_H(h)\, \mathrm{e}\left(-h\varphi^{-k}\right) = \frac{a_H(h)}{h}\left(1-\mathrm{e}\left(h\varphi^{-k+2}\right)\right).\end{aligned}$$

In order to treat the error terms, we use the nonnegativity of κ_H. For any $\beta \in \mathbb{R}$ we have

$$\begin{aligned}&\left|\sum_{u<F_{k-1}} \mathrm{e}(\vartheta Z(u))\, \kappa_H(p_k(n)-p_k(u)-\beta)\right|\\ &\qquad \le \frac{1}{H}\sum_{|h|\le H}\left(1-\frac{|h|}{H+1}\right)\sum_{u<F_{k-1}} \mathrm{e}\left((-1)^k h(n-u)\varphi - h\beta\right),\end{aligned}$$

where the right hand side is a nonnegative real number. Taking together E_1 and E_2 we get the estimate

$$|E_1+E_2| \le \frac{1}{H}\sum_{|h|\le H} c_H(h) \sum_{u<F_{k-1}} \mathrm{e}\left((-1)^k h(n-u)\varphi\right) \tag{5.8}$$

with

$$c_H(h) = \left(1-\frac{|h|}{H+1}\right)\left(\mathrm{e}\left(-h\varphi^{-k}\right)+\mathrm{e}\left(h\varphi^{-k+1}\right)\right).$$

We note the important fact that the right hand side of (5.8) is a nonnegative real number. Collecting the pieces, we obtain the statement of the proposition.

5.2.3 Proof of Theorem 5.2

The idea of proof is the same as in the proof of equation (4.18), but the Zeckendorf sum-of-digits function causes technical complications. More precisely, we have to use Proposition 5.4 instead of the much simpler inverse discrete Fourier transform. Moreover, instead of the Chinese Remainder Theorem, which we used for the estimation of $\left\|\ell_1 q_1^{-\lambda_1} + \ell_2 q_2^{-\lambda_2}\right\|$, we use the fact that the golden ratio φ has bounded partial quotients. It is well-known that such numbers are badly approximable, that is, we have $\|b\varphi\| > c/b$ for all b such that $b \neq 0$.

We begin with an application of van der Corput's inequality, which is Lemma 1.15. We obtain

$$\left|\sum_{n<N} \mathrm{e}\left(\vartheta s_q(n) + \beta Z(n)\right)\right|^2 \leq \frac{N+R}{R} \sum_{|r|<R} \left(1 - \frac{|r|}{R}\right) \sum_{0\leq n, n+r<N} \left(\vartheta\left(s_q(n+r) - s_q(n)\right) + \beta\left(Z(n+r) - Z(n)\right)\right).$$

We extend the sum-of-digits functions s_q and Z to $\mathbb{Z}$ by setting $s_q(n) = 0$ and $Z(n) = 0$ for $n < 0$. Omitting the condition $0 \leq n + r < N$ causes an error term $O(NR)$. We may therefore replace $N + R$ by N times a constant, since the left hand side is bounded by N^2. Moreover, we apply Lemma 1.17 and Lemma 1.29, which allow replacing s_q by $s_{q,\lambda}$ and Z by Z_k for the price of an additional error term $O(N^2R/F_{k-1} + N^2R/q^\lambda)$. Finally, let N' be the largest multiple of q^λ not greater than N. Collecting the error terms, we obtain

$$\left|\sum_{n<N} \mathrm{e}\left(\vartheta s_q(n) + \beta Z(n)\right)\right|^2 \ll O\left(NR + Nq^\lambda + N^2R/F_{k-1} + N^2R/q^\lambda\right) + S(N, r, \lambda, k), \quad (5.9)$$

where

$$S(N, r, \lambda, k) = \frac{N}{R} \sum_{|r|<R} \left(1 - \frac{|r|}{R}\right) \times \sum_{n<N'} \mathrm{e}\left(\vartheta s_{q,\lambda}(n+r)\right) \mathrm{e}\left(-\vartheta s_{q,\lambda}(n)\right) \mathrm{e}\left(\beta Z_k(n+r)\right) \mathrm{e}\left(-\beta Z_k(n)\right). \quad (5.10)$$

The first two factors are easy to handle using the discrete Fourier transform. We have

$$\mathrm{e}\left(\vartheta s_{q,\lambda}\right)(m) = \sum_{\ell<q^\lambda} \mathrm{e}\left(\ell m q^{-\lambda}\right) G_\lambda(\ell, \vartheta) \quad (5.11)$$

and

$$\mathrm{e}\left(-\vartheta s_{q,\lambda}\right)(m) = \sum_{\ell<q^\lambda} \mathrm{e}\left(\ell m q^{-\lambda}\right) \overline{G_\lambda(-\ell, \vartheta)}, \quad (5.12)$$

where

$$G_\lambda(\ell, \vartheta) = \frac{1}{q^\lambda} \sum_{u<q^\lambda} \mathrm{e}\left(\vartheta s_q(u) - \ell u q^{-\lambda}\right).$$

The third and the fourth factor in (5.10) are replaced using (5.2), which yields 16 summands. Each of these summands is a product of two expressions of the form (5.11) or (5.12), followed by a product of two summands from the right hand side of (5.2). We distinguish between three cases.

Case 1: Exactly one of the factors is an error term. Without loss of generality we assume that this is the second factor. The contribution of this case to the sum $S(N, r, \lambda, k)$ consists of a sum of four expressions of the form

$$\frac{N}{R} \sum_{|r|<R} \left(1 - \frac{|r|}{R}\right) \sum_{n<N} \sum_{\substack{\ell_1,\ell_2<q^\lambda \\ |h_1|\le H}} G_\lambda(\ell_1)\overline{G_\lambda(-\ell_2)} b_H^{(i)}(h_1) M_k^{(i)}(h_1,\beta) \\ \times \mathrm{e}\left(\ell_1(n+r)q^{-\lambda} + \ell_2 n q^{-\lambda} + h_1(-1)^k n\varphi\right) \\ \times O\left(\frac{1}{H} \sum_{|h_2|\le H} c_H^{(j)}(h_2)\, \mathrm{e}\left((-1)^k h_2 n\varphi\right) \sum_u \mathrm{e}\left(-(-1)^k h_2 u\varphi\right)\right) \quad (5.13)$$

where $0 \le u < F_{k-1}$ if $j = 1$ and $F_{k-1} \le u < F_k$ if $j = 2$. The expression in the error term is a nonnegative real number, as stated in Proposition 5.4. We estimate the Fourier coefficients $G_\lambda(\ell)$ and the sum over r trivially, moreover we use the estimate for $b_H^{(i)}(h_1)$ and obtain as the contribution to $S(N, r, \lambda, k)$

$$N\frac{q^{2\lambda}}{H} \sum_{|h_1|\le H} F_k \min\left\{\frac{1}{|h_1|}, \varphi^{-k+2}\right\} \sum_{|h_2|\le H} \left|\sum_u \mathrm{e}\left(-(-1)^k h_2 u\varphi\right)\right| \left|\sum_{n<N} \mathrm{e}\left((-1)^k h_2 n\varphi\right)\right| \\ \ll N\frac{q^{2\lambda}}{H} F_k \log H \sum_{|h|\le H} \min\left\{F_k, \|h\varphi\|^{-1}\right\} \min\left\{N, \|h\varphi\|^{-1}\right\} \\ \ll N\frac{q^{2\lambda} F_k \log H}{H} \sum_{|h|\le H} \min\left\{F_k N, \|h\varphi\|^{-2}\right\}.$$

We estimate the sum with the help of the following lemma.

Lemma 5.8. *Let I be a finite interval in $\mathbb{Z}$. Assume that K and t are real numbers and $K \ge 1$. Then*

$$\sum_{h\in I} \min\left\{K, \frac{1}{\|t + h\varphi\|^2}\right\} \ll \sqrt{K}\,|I| + K \log|I|.$$

Proof. We use the following well-known discrepancy estimate for the sequence $(n\varphi)_n$. We have

$$D_I(\varphi, t) = \sup_{0\le a<b\le 1} \big|\,|\{h \in I : a \le \{h\varphi + t\} < b\}| - (b-a)\,|I|\,\big| \ll \log|I|.$$

Let $M \ge 1$ be an integer satisfying $M^2 \le K$, which we chose later. We decompose the interval $[0, 1)$ into smaller intervals of length M^{-1} and set

$$A_\ell = |\{h : h \in I, \ell/M \le h\varphi + t < (\ell+1)/M\}|.$$

By the discrepancy estimate we have

$$A_\ell = \frac{|I|}{M} + O(\log |I|).$$

We obtain

$$\begin{aligned}\sum_{h\in I} \min\left\{K, \|t+h\varphi\|^{-2}\right\} &= \sum_{\ell<M} \sum_{\substack{h\in I\\ \ell/M\le\{h\varphi+t\}<(\ell+1)/M}} \min\left\{K, \|h\varphi\|^{-2}\right\}\\ &\le K(A_0+A_{M-1}) + \sum_{1\le\ell<M/2} (\ell/M)^{-2}A_\ell + \sum_{M/2\le\ell<M-1} (1-(\ell+1)/M)^{-2}A_\ell\\ &\ll \frac{K\,|I|}{M} + M^2\frac{|I|}{M} + K\log|I| + M^2\log|I| \ll \frac{K\,|I|}{M} + K\log|I|\,.\end{aligned}$$

In order to minimize this, we choose $M = \lfloor K^{1/2}\rfloor \ge 1$, which yields the statement of the lemma. □

The contribution to $S(N,r,\lambda,k)$ of the first case can therefore be estimated by

$$O\left(N^{3/2}q^{2\lambda}F_k^{3/2}\log H\frac{1}{H} + N^2q^{2\lambda}F_k^2(\log H)^2\frac{1}{H}\right). \tag{5.14}$$

Case 2: Both factors are error terms. This case is treated in a similar way. We obtain the contribution

$$\begin{aligned}&N\sum_{n<N}\sum_{\ell_1,\ell_2<q^\lambda} G_\lambda(\ell_1)\overline{G_\lambda(-\ell_2)}\,\mathrm{e}\left(\ell_1(n+r)q^{-\lambda} + \ell_2 nq^{-\lambda}\right)\\ &\times O\left(\frac{1}{H}\sum_{|h_1|\le H} c_H^{(i)}(h_1)\,\mathrm{e}\left((-1)^k h_1 n\varphi\right)\sum_u \mathrm{e}\left(-(-1)^k h_2 u\varphi\right)\right)\\ &\times O\left(\frac{1}{H}\sum_{|h_2|\le H} c_H^{(j)}(h_2)\,\mathrm{e}\left((-1)^k h_2 n\varphi\right)\sum_u \mathrm{e}\left(-(-1)^k h_2 u\varphi\right)\right).\end{aligned} \tag{5.15}$$

By estimating the first error term trivially by F_k and using Parseval's identity this can be bounded by

$$\begin{aligned}&\frac{N}{H}\left(\sum_{\ell<q^\lambda}|G_\lambda(\ell)|\right)^2 F_k\sum_{|h_2|\le H}\left|\sum_{n<N'}\mathrm{e}\left(h_2 n\varphi\right)\right|\left|\sum_u \mathrm{e}\left(h_2 u\varphi\right)\right|\\ &\ll N\frac{q^\lambda F_k}{H}\sum_{|h_2|\le H}\min\left\{F_k, \|h_2\varphi\|^{-1}\right\}\min\left\{N', \|h_2\varphi\|^{-1}\right\}\\ &\ll N\frac{q^\lambda F_k}{H}\sum_{|h_2|\le H}\min\left\{F_kN, \|h_2\varphi\|^{-2}\right\},\end{aligned} \tag{5.16}$$

which by Lemma 5.8 gives the contribution

$$O\left(N^{3/2}F_k^{3/2}q^\lambda + N^2F_k^2q^\lambda \log H\frac{1}{H}\right). \tag{5.17}$$

Case 3: In each of the two factors coming from the third and fourth terms in (5.10) we take one of the two main terms in (5.2). This contributes four summands. In this case we have to estimate

$$\begin{aligned} N \sum_{\substack{\ell_1,\ell_2<q^\lambda \\ |h_1|,|h_2|\le H}} & G_\lambda(\ell_1)\overline{G_\lambda(-\ell_2)}b_H^{(i)}(h_1)\overline{b_H^{(j)}(-h_2)}M_k^{(i)}(h_1,\beta)\overline{M_k^{(j)}(-h_2,\beta)} \\ & \times \frac{1}{R^2}\sum_{|r|<R}(R-|r|)\,\mathrm{e}\left(r\left(\frac{\ell_1}{q^\lambda}+h_1(-1)^k\varphi\right)\right)\sum_{n<N'}\mathrm{e}\left(n\left(\frac{\ell_1+\ell_2}{q^\lambda}+(h_1+h_2)(-1)^k\varphi\right)\right), \end{aligned} \tag{5.18}$$

where $i,j\in\{1,2\}$. We treat the case that $h_1+h_2=0$ first. By construction, we have $q^\lambda \mid N'$, therefore the case that $\ell_1+\ell_2\not\equiv 0 \bmod q^\lambda$ does not contribute anything. We assume therefore that $\ell_1+\ell_2\equiv 0 \bmod q^\lambda$.

By Lemma 1.31 we have the estimate $b_H^{(j)}M_k^{(j)}(h,\beta)\ll e^{-ck}$ for some $c>0$. Moreover, we use the identity

$$\sum_{|r|<R}(R-|r|)\,\mathrm{e}\,(rx)=\left|\sum_{r<R}\mathrm{e}\,(rx)\right|^2,$$

Parseval's identity and Lemma 5.8 to bound the contribution of the case that $h_1+h_2=0$ by

$$\begin{aligned} &\frac{N^2}{R^2}\sup_{h\in\mathbb{Z}}\left|b_H^{(j)}M_k^{(j)}(h,\beta)\right|\sum_{\ell_1<q^\lambda}|G_\lambda(\ell_1)|^2\sum_{|h_1|\le H}F_{k-1}\min\left\{\frac{1}{|h_1|},\varphi^{-k+2}\right\} \\ &\qquad\times\min\left\{R^2,\left\|\frac{l_1}{q^\lambda}+h_1(-1)^k\varphi\right\|^{-2}\right\} \\ &\qquad\ll \frac{N^2}{R^2}e^{-ck}F_k\sup_{\ell_1\in\mathbb{Z}}\sum_{|h|\le H}\min\left\{\frac{1}{|h|},\frac{1}{F_k}\right\}\min\left\{R^2,\left\|\ell_1/q^\lambda+|h|\,\varphi\right\|^{-2}\right\} \\ &\qquad\ll \frac{N^2}{R^2}e^{-ck}F_k\sum_{s<H/F_k}\min\left\{\frac{1}{sF_k},\frac{1}{F_k}\right\}\sup_{t\in\mathbb{R}}\sum_{h<F_k}\min\left\{R^2,\|t+(sF_k+h)\varphi\|^{-2}\right\} \\ &\qquad\ll \frac{N^2}{R^2}e^{-ck}\log H(RF_k+R^2\log F_k) \\ &\qquad\ll N^2e^{-ck}\log H\left(\frac{F_k}{R}+\log F_k\right). \end{aligned} \tag{5.19}$$

Now we assume that $h\neq 0$. We set $\ell=\ell_1+\ell_2$ and $h=h_1+h_2$. Since φ is badly approximable, there is a constant c such that

$$\left|hq^\lambda\varphi+\ell\right|>\frac{c}{q^\lambda h}$$

for all (ℓ, h) such that $h \neq 0$. Dividing by q^λ, we obtain

$$\left| h\varphi + \frac{\ell}{q^\lambda} \right| > \frac{c}{q^{2\lambda} h}$$

for all ℓ and h, in particular

$$\left\| h\varphi + \frac{\ell}{q^\lambda} \right\| > c\frac{1}{q^{2\lambda} h}. \tag{5.20}$$

We estimate the Fourier terms in (5.18) and the sum over r trivially, moreover we replace the sum over ℓ_1 and ℓ_2 by a sup. It remains to estimate the expression

$$\begin{aligned} Nq^{2\lambda} \sum_{\substack{|h_1|,|h_2|\le H \\ h_1+h_2\neq 0}} \sup_{\ell\in\mathbb{Z}} \left| \sum_{n<N'} \mathrm{e}\left(n\left(\frac{\ell}{q^\lambda} + (h_1+h_2)(-1)^k\varphi \right) \right) \right| \\ = Nq^{2\lambda} \sum_{1\le|h|\le 2H} (2H+1-|h|) \sup_{\ell\in\mathbb{Z}} \left| \sum_{n<N'} \mathrm{e}\left(n\left(\frac{\ell}{q^\lambda} + h\varphi \right) \right) \right| \\ \ll NHq^{2\lambda} \sum_{1\le|h|\le 2H} Hq^{2\lambda} \ll NH^3 q^{4\lambda}. \end{aligned} \tag{5.21}$$

We collect the error terms from (5.9), (5.14), (5.17), (5.19) and (5.21) and obtain for $\beta \notin \mathbb{Z}$, $R, H, \lambda \ge 1$ and $k \ge 2$

$$\begin{aligned} &\left| \sum_{n<N} \mathrm{e}\left(\vartheta s_q(n) + \beta Z(n)\right) \right|^2 \\ &\quad \ll N^2\frac{R}{q^\lambda} + N^2\frac{R}{F_k} + N^2 e^{-ck}\log H\frac{F_k}{R} + N^2 q^{2\lambda}F_k^2(\log H)^2\frac{1}{H} + N^2F_k^2q^\lambda\log H\frac{1}{H} \\ &\quad + N^2e^{-ck}\log H\log F_k + N^{3/2}q^{2\lambda}F_k^{3/2}\log H\frac{1}{H} + N^{3/2}F_k^{3/2}q^\lambda + NR + Nq^\lambda + NH^3q^{4\lambda}. \end{aligned}$$

It remains to choose k, λ, R and H. We introduce a new variable a and choose

$$\lambda = \left\lfloor a\frac{\log N}{\log q} \right\rfloor, \quad k = \left\lfloor a\frac{\log N}{\log \varphi} \right\rfloor, \quad R = \left\lfloor N^{a-ca/(2\log\varphi)} \right\rfloor, \quad H = \left\lfloor N^{4a-ca/2\log\varphi)} \right\rfloor.$$

The rest of the calculation is straightforward, yielding a contribution $\ll N^{1-\eta}$ for each of the eleven error terms, if a is chosen small enough. This finishes the proof of Theorem 5.2.

We also note that an admissible value of the exponent η could easily be obtained as soon as an admissible value for c is found (which can be done by studying the proof of Lemma 1.31). We skip the details.

Bibliography

[1] R. C. Baker, W. D. Banks, J. Brüdern, I. E. Shparlinski, and A. J. Weingartner, *Piatetski-Shapiro sequences*, Acta Arith., 157 (2013), pp. 37–68.

[2] W. D. Banks and I. E. Shparlinski, *Prime numbers with Beatty sequences*, Colloq. Math., 115 (2009), pp. 147–157.

[3] G. Barat and P. J. Grabner, *Distribution of binomial coefficients and digital functions*, J. London Math. Soc. (2), 64 (2001), pp. 523–547.

[4] S. Beatty, *Problem 3173*, Amer. Math. Monthly, 33 (1926).

[5] R. Bellman and H. N. Shapiro, *On a problem in additive number theory*, Ann. of Math. (2), 49 (1948), pp. 333–340.

[6] J.-P. Bertrandias, *Suites pseudo-aléatoires et critères d'équirépartition modulo un*, Compositio Math., 16 (1964), pp. 23–28 (1964).

[7] J. Bésineau, *Indépendance statistique d'ensembles liés à la fonction "somme des chiffres"*, Acta Arith., 20 (1972), pp. 401–416.

[8] X. D. Cao and W. G. Zhai, *Distribution of square-free numbers of the form* $[n^c]$. *II*, Acta Math. Sinica (Chin. Ser.), 51 (2008), pp. 1187–1194.

[9] J. Coquet, *Sur les fonctions q-multiplicatives pseudo-aléatoires*, C. R. Acad. Sci. Paris Sér. A-B, 282 (1976), pp. Ai, A175–A178.

[10] ——, *Contribution à l'étude harmonique des suites arithmétiques*, Thèse d'Etat, Orsay, 1978.

[11] J. Coquet, T. Kamae, and M. Mendès France, *Sur la mesure spectrale de certaines suites arithmétiques*, Bull. Soc. Math. France, 105 (1977), pp. 369–384.

[12] J. Coquet and M. Mendès France, *Suites à spectre vide et suites pseudo-aléatoires*, Acta Arith., 32 (1977), pp. 99–106.

[13] J. Coquet, G. Rhin, and P. Toffin, *Représentations des entiers naturels et indépendance statistique. II*, Ann. Inst. Fourier (Grenoble), 31 (1981), pp. ix, 1–15.

[14] T. W. Cusick, 2012. private communication.

[15] T. W. Cusick, Y. Li, and P. Stănică, *On a combinatorial conjecture*, Integers, 11 (2011), pp. A17, 17.

[16] H. Delange, *Sur les fonctions q-additives ou q -multiplicatives*, Acta Arith., 21 (1972), pp. 285–298. (errata insert).

[17] J.-M. Deshouillers, *Sur la répartition des nombres $[n^c]$ dans les progressions arithmétiques*, C. R. Acad. Sci. Paris Sér. A-B, 277 (1973), pp. A647–A650.

[18] J.-M. Deshouillers, M. Drmota, and J. F. Morgenbesser, *Subsequences of automatic sequences indexed by $\lfloor n^c \rfloor$ and correlations*, J. Number Theory, 132 (2012), pp. 1837–1866.

[19] E. Dijkstra, *problem section, problem 563*, Nieuw Archief voor Wiskunde, XXVII (1980), p. 115.

[20] E. W. Dijkstra, *Selected writings on computing: a personal perspective*, Texts and Monographs in Computer Science, Springer-Verlag, New York, 1982. Including a paper co-authored by C. S. Scholten.

[21] M. Drmota and P. J. Grabner, *Analysis of digital functions and applications*, in Combinatorics, automata and number theory, vol. 135 of Encyclopedia Math. Appl., Cambridge Univ. Press, Cambridge, 2010, pp. 452–504.

[22] P. Flajolet, P. Grabner, P. Kirschenhofer, H. Prodinger, and R. F. Tichy, *Mellin transforms and asymptotics: digital sums*, Theoret. Comput. Sci., 123 (1994), pp. 291–314.

[23] P. Flajolet and A. Odlyzko, *Singularity analysis of generating functions*, SIAM J. Discrete Math., 3 (1990), pp. 216–240.

[24] P. Flajolet and R. Sedgewick, *Analytic combinatorics*, Cambridge University Press, Cambridge, 2009.

[25] J.-P. Flori and H. Randriam, *On the number of carries occurring in an addition mod $2^k - 1$*, Integers, 12 (2012), pp. 601–647.

[26] J.-P. Flori, H. Randriam, G. Cohen, and S. Mesnager, *On a conjecture about binary strings distribution*, in Sequences and their applications—SETA 2010, vol. 6338 of Lecture Notes in Comput. Sci., Springer, Berlin, 2010, pp. 346–358.

[27] E. Fouvry and C. Mauduit, *Sommes des chiffres et nombres presque premiers*, Math. Ann., 305 (1996), pp. 571–599.

[28] H. Furstenberg, *Algebraic functions over finite fields*, J. Algebra, 7 (1967), pp. 271–277.

[29] A. O. Gel′fond, *Sur les nombres qui ont des propriétés additives et multiplicatives données*, Acta Arith., 13 (1967/1968), pp. 259–265.

[30] S. W. Graham and G. Kolesnik, *van der Corput's method of exponential sums*, vol. 126 of London Mathematical Society Lecture Note Series, Cambridge University Press, Cambridge, 1991.

[31] M. L. J. Hautus and D. A. Klarner, *The diagonal of a double power series*, Duke Math. J., 38 (1971), pp. 229–235.

[32] E. Hlawka and J. Schoissengeier, *Zahlentheorie*, Manzsche Verlags- und Universitätsbuchhandlung, Vienna, 1979. Eine Einführung, Vorlesungen über Mathematik.

[33] D.-H. Kim, *On the joint distribution of q-additive functions in residue classes*, J. Number Theory, 74 (1999), pp. 307–336.

[34] ——, *On the distribution modulo 1 of q-additive functions*, Acta Math. Hungar., 90 (2001), pp. 75–83.

[35] E. E. Kummer, *über die ergänzungssätze zu den allgemeinen reciprocitätsgesetzen*, J. Reine Angew. Math., 44 (1852), pp. 93–146.

[36] D. Leitmann and D. Wolke, *Primzahlen der Gestalt* $[f(n)]$, Math. Z., 145 (1975), pp. 81–92.

[37] K. Mahler, *The spectrum of an array and its application to the study of the translation properties of a simple class of arithmetical functions II. On the translation properties of a simple class of arithmetical functions*, J. Math. and Physics, 6 (1927), pp. 158–163.

[38] C. Mauduit, *Multiplicative properties of the Thue-Morse sequence*, Period. Math. Hungar., 43 (2001), pp. 137–153.

[39] C. Mauduit and J. Rivat, *Répartition des fonctions q-multiplicatives dans la suite* $([n^c])_{n\in\mathbf{N}},\ c>1$, Acta Arith., 71 (1995), pp. 171–179.

[40] ——, *Propriétés q-multiplicatives de la suite* $\lfloor n^c\rfloor,\ c>1$, Acta Arith., 118 (2005), pp. 187–203.

[41] ——, *La somme des chiffres des carrés*, Acta Math., 203 (2009), pp. 107–148.

[42] ——, *Sur un problème de Gelfond: la somme des chiffres des nombres premiers*, Ann. of Math. (2), 171 (2010), pp. 1591–1646.

[43] C. Mauduit, J. Rivat, and A. Sárközy, *On the pseudo-random properties of* n^c, Illinois J. Math., 46 (2002), pp. 185–197.

[44] M. Mendès France, *Nombres normaux. Applications aux fonctions pseudo-aléatoires*, J. Analyse Math., 20 (1967), pp. 1–56.

[45] H. L. Montgomery, *The analytic principle of the large sieve*, Bull. Amer. Math. Soc., 84 (1978), pp. 547–567.

[46] J. F. Morgenbesser, *The sum of digits of* $\lfloor n^c\rfloor$, Acta Arith., 148 (2011), pp. 367–393.

[47] J. F. Morgenbesser and L. Spiegelhofer, *A reverse order property of correlation measures of the sum-of-digits function*, Integers, 12 (2012), pp. Paper No. A47, 5.

[48] S. Northshield, *Stern's diatomic sequence* $0, 1, 1, 2, 1, 3, 2, 3, 1, 4, \ldots$, Amer. Math. Monthly, 117 (2010), pp. 581–598.

[49] R. Pemantle and M. C. Wilson, *Analytic combinatorics in several variables*, vol. 140 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 2013.

[50] O. Perron, *Die Lehre von den Kettenbrüchen. Bd I. Elementare Kettenbrüche*, B. G. Teubner Verlagsgesellschaft, Stuttgart, 1954. 3te Aufl.

[51] I. I. Pyateckiĭ-Šapiro, *On the distribution of prime numbers in sequences of the form* $[f(n)]$, Mat. Sbornik N.S., 33(75) (1953), pp. 559–566.

[52] J. S. Rayleigh, *The Theory of sound*, vol. 1, Macmillan and Co., London, second ed., 1894.

[53] J. Rivat and P. Sargos, *Nombres premiers de la forme* $\lfloor n^c \rfloor$, Canad. J. Math., 53 (2001), pp. 414–433.

[54] Z. Tu and Y. Deng, *A conjecture about binary strings and its applications on constructing Boolean functions with optimal algebraic immunity*, Des. Codes Cryptogr., 60 (2011), pp. 1–14.

[55] I. Urbiha, *Some properties of a function studied by de Rham, Carlitz and Dijkstra and its relation to the (Eisenstein-)Stern's diatomic sequence*, Math. Commun., 6 (2001), pp. 181–198.

[56] Ś. Zą bek, *Sur la périodicité modulo m des suites de nombres* $\binom{n}{k}$, Ann. Univ. Mariae Curie-Sklodowska. Sect. A, 10 (1956), pp. 37–47 (1958).

[57] E. Zeckendorf, *Répresentation des nombres naturels par une somme de nombres de Fibonacci ou de nombres de Lucas*, Bull. Soc. Roy. Sci. Liège, 41 (1972), pp. 179–182.

Acknowledgements

First I want to thank my advisors, Michael Drmota and Joël Rivat. They always had time for my questions at short notice and helped me to improve my understanding of mathematical questions and beyond.

Special thanks go to my colleagues Zbigniew Golebiewski, Benoît Loridant, Johannes Morgenbesser, Thomas Stoll, Michael Wallner and Daniel Weller, who showed interest in the progress of my studies and provided helpful advice.

I thank Cécile Dartyge and Gerhard Larcher for reviewing my thesis and Peter Grabner for accepting to chair the final exam.

Moreover, I thank the Institute for Discrete Mathematics and Geometry at the TU Wien and the Institute de Mathématiques de Luminy at the Université d'Aix-Marseille for providing good working conditions, and the Austrian Science Foundation FWF for supporting me financially for the course of my studies.

Finally I want to thank Johanna Jandl, who supported the progress of my studies in many ways.