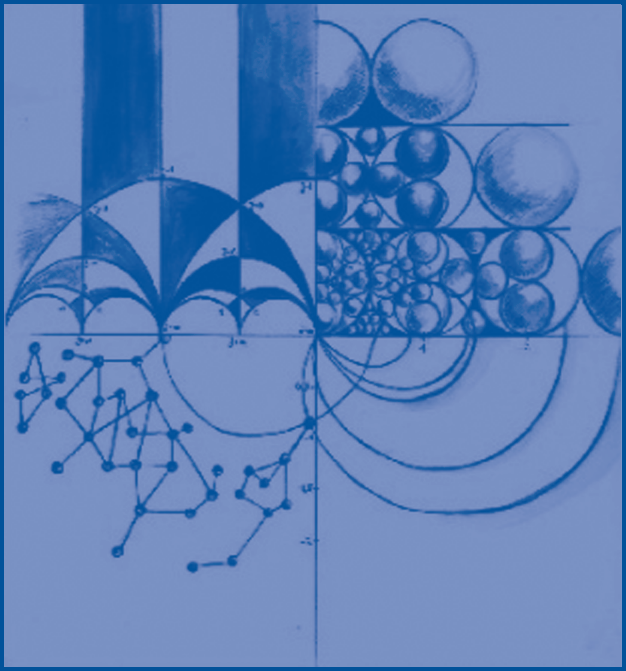
EMS Series of Lectures in Mathematics

Kathrin Bringmann
Yann Bugeaud
Titus Hilberdink
Jürgen Sander

Four Faces of Number Theory



European Mathematical Society

EMS Series of Lectures in Mathematics
Edited by Andrew Ranicki (University of Edinburgh, U.K.)

EMS Series of Lectures in Mathematics is a book series aimed at students, professional mathematicians and scientists. It publishes polished notes arising from seminars or lecture series in all fields of pure and applied mathematics, including the reissue of classic texts of continuing interest. The individual volumes are intended to give a rapid and accessible introduction into their particular subject, guiding the audience to topics of current research and the more advanced and specialized literature.

Previously published in this series:
Katrin Wehrheim, *Uhlenbeck Compactness*
Torsten Ekedahl, *One Semester of Elliptic Curves*
Sergey V. Matveev, *Lectures on Algebraic Topology*
Joseph C. Várilly, *An Introduction to Noncommutative Geometry*
Reto Müller, *Differential Harnack Inequalities and the Ricci Flow*
Eustasio del Barrio, Paul Deheuvels and Sara van de Geer, *Lectures on Empirical Processes*
Iskander A. Taimanov, *Lectures on Differential Geometry*
Martin J. Mohlenkamp and María Cristina Pereyra, *Wavelets, Their Friends, and What They Can Do for You*
Stanley E. Payne and Joseph A. Thas, *Finite Generalized Quadrangles*
Masoud Khalkhali, *Basic Noncommutative Geometry*
Helge Holden, Kenneth H. Karlsen, Knut-Andreas Lie and Nils Henrik Risebro, *Splitting Methods for Partial Differential Equations with Rough Solutions*
Koichiro Harada, *"Moonshine" of Finite Groups*
Yurii A. Neretin, *Lectures on Gaussian Integral Operators and Classical Groups*
Damien Calaque and Carlo A. Rossi, *Lectures on Duflo Isomorphisms in Lie Algebra and Complex Geometry*
Claudio Carmeli, Lauren Caston and Rita Fioresi, *Mathematical Foundations of Supersymmetry*
Hans Triebel, *Faber Systems and Their Use in Sampling, Discrepancy, Numerical Integration*
Koen Thas, *A Course on Elation Quadrangles*
Benoît Grébert and Thomas Kappeler, *The Defocusing NLS Equation and Its Normal Form*
Armen Sergeev, *Lectures on Universal Teichmüller Space*
Matthias Aschenbrenner, Stefan Friedl and Henry Wilton, *3-Manifold Groups*
Hans Triebel, *Tempered Homogeneous Function Spaces*

Kathrin Bringmann
Yann Bugeaud
Titus Hilberdink
Jürgen Sander

Four Faces of Number Theory

European Mathematical Society

Authors:

Kathrin Bringmann
Mathematisches Institut
Universität zu Köln
Gyrhofstr. 8b
50931 Köln
Germany

E-mail: kbringma@math.uni-koeln.de

Yann Bugeaud
Institut de Recherche Mathématique Avancée, UMR 7501
Université de Strasbourg et CNRS
7, rue René Descartes
67084 Strasbourg
France

E-mail: yann.bugeaud@math.unistra.fr

Titus Hilberdink
Department of Mathematics and Statistics
University of Reading
Whiteknights P.O. Box 220
Reading RG6 6AX
UK

E-mail: t.w.hilberdink@reading.ac.uk

Jürgen Sander
Institut für Mathematik und Angewandte Informatik
Universität Hildesheim
Samelsonplatz 1
31141 Hildesheim
Germany

E-mail: sander@imai.uni-hildesheim.de

2010 Mathematics Subject Classification: Primary: 11-02, 11J81; Secondary 11A63, 11J04, 11J13, 11J68, 11J70, 11J87, 68R15

Key words: Transcendence, algebraic number, Schmidt Subspace Theorem, continued fraction, digital expansion, Diophantine approximation

ISBN 978-3-03719-142-2

The Swiss National Library lists this publication in The Swiss Book, the Swiss national bibliography, and the detailed bibliographic data are available on the Internet at http://www.helveticat.ch.

Contact address:
European Mathematical Society Publishing House
Seminar for Applied Mathematics
ETH-Zentrum SEW A27
CH-8092 Zürich
Switzerland

Phone: +41 (0)44 632 34 36
Email: info@ems-ph.org
Homepage: www.ems-ph.org

Typeset using the authors' $\TeX$ files: le-tex publishing services GmbH, Leipzig, Germany
Printing and binding: Beltz Bad Langensalza GmbH, Bad Langensalza, Germany
∞ Printed on acid free paper
9 8 7 6 5 4 3 2 1

Foreword

Number theory is a fascinating branch of mathematics. There are many examples of number theoretical problems that are easy to understand but difficult to solve, often enough only by use of advanced methods from other mathematical disciplines. This might be one of the reasons why number theory is considered to be such an attractive field with many connections to other areas of mathematical research.

In August 2012 the number theory group of the Department of Mathematics at Würzburg University, under the aegis of Jörn Steuding, organized an international summer school entitled Four Faces of Number Theory. In the frame of this event about fifty participants, mostly PhD students from all over the world, but also a few local participants, even undergraduate students, learned in four courses about different aspects of modern number theory. These courses highlight a strong interplay between number theory and other fields like combinatorics, functional analysis and graph theory. They will be of interest to (under)graduate students aiming to discover various aspects of number theory and their relationship with other areas of mathematics.

Kathrin Bringmann from Cologne gave an introduction to the theory of modular forms and, in particular, so-called Mock theta-functions, a topic which had been untouched for decades but has obtained much attention during the last five years.

Yann Bugeaud from Strasbourg lectured about expansions of algebraic numbers. Despite some recent progress, presented in his essay, questions like ‘does the digit 7 occur infinitely often in the decimal expansion of square root of two?’ remain very far from being answered. Here combinatorics on words and transcendence theory are combined to derive new information on the sequence of decimals of algebraic numbers and on their continued fraction expansions.

Titus Hilberdink from Reading lectured about a recent and rather unexpected approach to extreme values of the Riemann zeta-function by use of (multiplicative) Toeplitz matrices and functional analysis.

Finally, Jürgen Sander from Hildesheim gave an introduction to algebraic graph theory and the impact of number theoretical methods on fundamental questions about the spectra of graphs and the analogue of the Riemann hypothesis.

In this volume the reader can find the course notes from this summer school and in some places further additional material. Each of these courses is essentially self-contained (although a background in number theory and analysis might be useful). In all four courses recent research results are included indicating how easily one can approach frontiers of current research in number theory by elementary and basic analytic methods.

The picture on the front page shows the poster created by Nicola Oswald from Würzburg University for the summer school. The editing of the course notes had been done by Rasa Steuding from Würzburg University. The authors are most grateful to both of them for their help. Last but not least, we sincerely thank Jörn Steuding from Würzburg University for initiating the summer school and the publication of these notes as well as for his editorial work.

The authors, March 2014

Contents

Chapter 1

Asymptotic formulas for modular forms and related functions

Kathrin Bringmann

Contents

1 Introduction

In this paper[1], we aim to describe how “modularity” can be useful for studying the asymptotic behavior of arithmetically interesting functions. We do not attempt to present all that is known, but rather work with examples to give an idea about basic concepts.

Let us recall the objects of interest. In the words of Barry Mazur,

> “Modular forms are functions on the complex plane that are inordinately symmetric. They satisfy so many symmetries that their mere existence seem like accidents. But they do exist.”

The modular forms alluded to in this quote are meromorphic functions on the complex upper half-plane $\mathbb{H} := \{\tau \in \mathbb{C}; \mathrm{Im}(\tau) > 0\}$ that satisfy (if f is modular of weight $k \in \mathbb{Z}$ for $\mathrm{SL}_2(\mathbb{Z})$)

$$f\left(\frac{a\tau+b}{c\tau+d}\right) = (c\tau+d)^k f(\tau)$$
$$\forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) := \left\{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{Mat}_2(\mathbb{Z}); ad-bc=1\right\}. \tag{1.1}$$

Moreover, as a technical growth condition, one requires the function to be “meromorphic at the cusps”. Note that the transformation law can be generalized to subgroups, to include multipliers or half-integral weight. We do not state the specific form of

[1] The research of the author was supported by the Alfried Krupp Prize for Young University Teachers of the Krupp Foundation.

these transformations here, but we will later treat special cases, for example in Theorem 3.1.

An important property of modular forms is that they have Fourier expansions of the form $f(\tau) = \sum_{n\in\mathbb{Z}} a(n)e^{2\pi i n\tau}$. This follows from the transformation law (1.1), the fact that $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, and the meromorphicity on $\mathbb{H}$. Meromorphicity at the cusps now says that $a(n) \neq 0$ for only finitely many $n < 0$. The coefficients $a(n)$ often encode interesting arithmetic information, such as the number of representations of n by a (positive definite) quadratic form, just to give one of the numerous examples.

Modular forms play an important role in many areas like physics, representation theory, the theory of elliptic curves (in particular the proof of Fermat's Last Theorem), quadratic forms, and partitions, just to mention a few. Establishing modularity is of importance because it provides powerful machineries which can be employed to prove important results. For example, identities may be reduced to a finite calculation of Fourier coefficients (Sturm's Theorem), asymptotic formulas for Fourier coefficients can be obtained by Tauberian Theorems or the Circle Method, and congruences may be proven by employing Serre's theory of p-adic modular forms. In this note we are particularly interested in asymptotic and exact formulas for Fourier coefficients of various modular q-series. In Section 2 we consider holomorphic modular forms, then in Section 3 allow growth in the cusps, in Section 4 turn to mock modular forms, and finally treat mixed mock modular forms in Section 5.

2 Classical modular forms

In this section we restrict, for simplicity, to forms of even integral weight k for $\mathrm{SL}_2(\mathbb{Z})$. For basic facts on modular forms and most of the details skipped in this section, we refer the reader to [31].

We call a modular form a *holomorphic modular form* if it is holomorphic on the upper half-plane and bounded at ∞. The space of holomorphic modular forms of weight k is denoted by M_k. If a holomorphic modular form exponentially decays towards ∞ it is called a *cusp form*. The associated space is denoted by S_k. Special holomorphic modular forms that are not cusp forms are given by the classical Eisenstein series and they have very simple explicit Fourier coefficients.

Definition. Formally define for $k \in \mathbb{N}$ the *Eisenstein series*

$$G_k(\tau) := \sideset{}{'}\sum_{m,n\in\mathbb{Z}} (m\tau + n)^{-k},$$

where the sum runs through all $(m,n) \in \mathbb{Z}^2 \setminus \{(0,0)\}$. We note that

$$G_{2k+1}(\tau) \equiv 0.$$

Theorem 2.1. *For $k \geq 4$ even, we have that $G_k \in M_k$.*

Proof. (sketch)
Step 1: Prove compact convergence.
Step 2: Apply modular transformations and reorder. □

Remark. We also require a normalized version of the Eisenstein series. For this, set $\Gamma_\infty := \left\{\left(\begin{smallmatrix}1 & n\\ 0 & 1\end{smallmatrix}\right); n \in \mathbb{Z}\right\}$ and for $M = \left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R}), k \in \mathbb{Z}$, and $f : \mathbb{H} \to \mathbb{C}$ we define the *Petersson slash operator* by

$$f|_k M(\tau) := (c\tau + d)^{-k} f\left(\frac{a\tau + b}{c\tau + d}\right).$$

Define for $k \geq 4$ an even integer

$$E_k(\tau) := \frac{1}{2} \sum_{M \in \Gamma_\infty \backslash \mathrm{SL}_2(\mathbb{Z})} 1|_k M(\tau),$$

where the sum runs through a complete set of representatives of right cosets of Γ_∞ in $\mathrm{SL}_2(\mathbb{Z})$. We have that

$$E_k(\tau) = \frac{1}{2\zeta(k)} G_k(\tau),$$

where $\zeta(s) := \sum_{n \geq 1} \frac{1}{n^s}$ ($\mathrm{Re}(s) > 1$) denotes the *Riemann zeta function*. Indeed, it is not hard to see that a set of representatives of $\Gamma_\infty \setminus \mathrm{SL}_2(\mathbb{Z})$ may be given by

$$\left\{\begin{pmatrix}\star & \star\\ c & d\end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}); (c, d) = 1\right\}.$$

Thus

$$E_k(\tau) = \frac{1}{2} \sum_{(c,d)=1} (c\tau + d)^{-k} = \frac{1}{2\zeta(k)} \sum_{r \geq 1} r^{-k} \sum_{(c,d)=1} (c\tau + d)^{-k}$$
$$= \frac{1}{2\zeta(k)} \sum_{\substack{(c,d)=1\\ r \in \mathbb{N}}} (cr\tau + dr)^{-k}.$$

Now (cr, dr) runs through $\mathbb{Z}^2 \setminus \{(0, 0)\}$ if r runs through $\mathbb{N}$ and (c, d) through $\mathbb{Z}^2$ under the restriction that $(c, d) = 1$. This gives the claim.

We next turn to computing the Fourier expansion of the Eisenstein series.

Theorem 2.2. *We have the Fourier expansion for $k \geq 4$ even*

$$E_k(\tau) = 1 - \frac{2k}{B_k} \sum_{n \geq 1} \sigma_{k-1}(n) e^{2\pi i n \tau},$$

where $\sigma_\ell(n) := \sum_{d|n} d^\ell$ and B_k is the kth Bernoulli number, given by the generating function

$$\frac{x}{e^x - 1} = \sum_{n \geq 0} B_n \frac{x^n}{n!}.$$

Proof. We instead compute the Fourier expansion of G_k and then use that for k even

$$\zeta(k) = (-1)^{\frac{k}{2}+1} \frac{(2\pi)^k B_k}{2k!}.$$

Due to the absolute convergence of the Eisenstein series we may reorder

$$\begin{aligned} G_k(\tau) &= {\sum_{m,n\in\mathbb{Z}}}' (m\tau+n)^{-k} = \sum_{n\neq 0} n^{-k} + \sum_{m\neq 0}\sum_{n\in\mathbb{Z}} (m\tau+n)^{-k} \\ &= 2\zeta(k) + 2\sum_{m>0}\sum_{n\in\mathbb{Z}} (m\tau+n)^{-k}. \end{aligned} \tag{1.2}$$

Now it is well-known by the Lipschitz summation formula (cf. pp. 65–72 of [30]) that

$$\sum_{n\in\mathbb{Z}} (\tau+n)^{-k} = \frac{(-2\pi i)^k}{(k-1)!} \sum_{n\geq 1} n^{k-1} e^{2\pi i n\tau}.$$

This gives that the second term in (1.2) equals

$$\frac{2(2\pi i)^k}{(k-1)!} \sum_{m\geq 1}\sum_{d\geq 1} d^{k-1} e^{2\pi i m d\tau} = \frac{2(2\pi i)^k}{(k-1)!} \sum_{m\geq 1}\sum_{d|m} d^{k-1} e^{2\pi i m\tau},$$

which gives the claim. □

Examples. We have the following special cases:

$$\begin{aligned} E_4(\tau) &= 1 + 240\sum_{n\geq 1} \sigma_3(n) e^{2\pi i n\tau}, \\ E_6(\tau) &= 1 - 504\sum_{n\geq 1} \sigma_5(n) e^{2\pi i n\tau}, \\ E_8(\tau) &= 1 + 480\sum_{n\geq 1} \sigma_7(n) e^{2\pi i n\tau}. \end{aligned}$$

We next turn to bounding Fourier coefficients of holomorphic modular forms. For this we split the space M_k into an Eisenstein series part and a cuspidal part.

Theorem 2.3. *Assume that $k \geq 4$ is even. Then*

$$M_k = \mathbb{C}E_k \oplus S_k.$$

Proof. Assume $f(\tau) = \sum_{n\geq 0} a(n) e^{2\pi i n\tau} \in M_k$. Then

$$f - a(0)E_k \in S_k.$$

□

Since the coefficients of the Eisenstein series part were given explicitly in Theorem 2.2, we next turn to bounding coefficients of cups forms.

Theorem 2.4 (Hecke bound). *For $f(\tau) = \sum_{n\geq 1} a(n)e^{2\pi i n\tau} \in S_k$ with $k > 0$, we have*

$$a(n) = O\left(n^{\frac{k}{2}}\right).$$

Proof. It is not hard to see that the function

$$\widetilde{f}(\tau) := \mathrm{Im}(\tau)^{\frac{k}{2}} |f(\tau)|$$

is bounded on $\mathbb{H}$. Indeed, one can show that $\widetilde{f}$ is invariant under $\mathrm{SL}_2(\mathbb{Z})$ and one may then bound $\widetilde{f}$ for sufficiently large imaginary part. Using Cauchy's Theorem, we can write for $n \geq 1$

$$a(n) = e^{2\pi n y} \int_0^1 f(x+iy)e^{-2\pi i n x} dx,$$

where $y > 0$ can be chosen arbitrary. This yields that

$$|a(n)| \leq y^{-\frac{k}{2}} e^{2\pi n y} \int_0^1 \widetilde{f}(x+iy)dx \leq c y^{-\frac{k}{2}} e^{2\pi n y}$$

for some constant $c > 0$ (independent of y). Picking $y = \frac{1}{n}$ gives

$$|a(n)| \leq c n^{\frac{k}{2}} e^{2\pi} = O\left(n^{\frac{k}{2}}\right). \qquad \square$$

Remark. The Ramanujan-Petersson Conjecture predicts that for $f \in S_k$ we have for any $\varepsilon > 0$ the (optimal) estimate

$$a(n) \ll_{\varepsilon, f} n^{\frac{k-1}{2}+\varepsilon}.$$

For integral weight $k \geq 2$ this conjecture was proven by Deligne using highly advanced techniques from algebraic geometry. His method begins with the fact that the vector space of cusp forms has a basis of common eigenfunctions of the so-called Hecke algebra and that the Fourier coefficients of these forms coincide with the eigenvalues.

Corollary 2.5. *Assume $k \geq 4$ is even and $f(\tau) = \sum_{n\geq 0} a(n)e^{2\pi i n\tau} \in M_k$. Then*

$$a(n) = -a(0)\frac{2k}{B_k}\sigma_{k-1}(n) + O\left(n^{\frac{k}{2}}\right).$$

Proof. The claim follows directly by combining Theorem 2.3 and Theorem 2.4. □

Remark. We have that

$$n^r \leq \sigma_r(n) = \sum_{d|n} d^r = n^r \sum_{d|n} d^{-r} \leq n^r \sum_{d\geq 1} d^{-r} = \zeta(r)n^r.$$

In particular, we have for $k \geq 4$ that for $f \in M_k$

$$a(n) = O\left(n^{k-1}\right)$$

and if $f \notin S_k$, this bound cannot be improved.

We next turn to defining explicit cusp forms whose construction generalizes that of Eisenstein series.

Definition. For $k \in \mathbb{N}$ and $n \in \mathbb{N}_0$, formally define the Poincaré series

$$P_{k,n}(\tau) := \sum_{M \in \Gamma_\infty \backslash \mathrm{SL}_2(\mathbb{Z})} e^{2\pi i n\tau}\big|_k M.$$

Note that

$$P_{k,0}(\tau) = 2E_k(\tau).$$

A calculation similar to the proof of Theorem 2.1 establishes the modularity of these functions.

Theorem 2.6. *For $k \geq 3$ the Poincaré series $P_{k,n}$ converges uniformly on every vertical strip in $\mathbb{H}$. In particular, $P_{k,n} \in M_k$, and for $n > 0$ we have $P_{k,n} \in S_k$.*

Poincaré series turn out to be useful, as they have explicitly computable Fourier expansions and integrating cusp forms against them recovers the Fourier coefficients of these forms.

To precisely state this result, we need the Petersson inner product. To define this, note that for $f, g \in M_k$

$$d\mu := \frac{dxdy}{y^2},$$

$$f(\tau)\overline{g(\tau)}y^k$$

are invariant under $\mathrm{SL}_2(\mathbb{Z})$.

Definition. A subset $\mathcal{F} \subset \mathbb{H}$ is called *a fundamental domain* of $\mathrm{SL}_2(\mathbb{Z})$ if the following hold:

(i) $\mathcal{F}$ is closed.

(ii) For every $\tau \in \mathbb{H}$ there exists $M \in \mathrm{SL}_2(\mathbb{Z})$ such that $M\tau \in \mathcal{F}$.

(iii) If τ and $M\tau$ ($M \in \mathrm{SL}_2(\mathbb{Z})$) are in the interior of $\mathcal{F}$, then $M = \pm\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$.

We also require the following explicit fundamental domain.

Theorem 2.7. *The following is a fundamental domain of* $\mathrm{SL}_2(\mathbb{Z})$*:*

$$\mathcal{F} := \left\{\tau \in \mathbb{H}; |\tau| \geq 1, |\mathrm{Re}(\tau)| \leq \frac{1}{2}\right\}.$$

Remark. Note that if $\tau = x + iy \in \mathcal{F}$, then $y \geq \frac{\sqrt{3}}{2}$.

Definition. For $f \in M_k$ and $g \in S_k$, formally define the *Petersson inner product* by

$$\langle f, g\rangle := \int_{\mathcal{F}} f(\tau)\overline{g(\tau)}y^k d\mu. \tag{1.3}$$

Later we also require a regularized version of this inner product. To introduce it, denote for $T > 0$

$$\mathcal{F}_T := \left\{\tau \in \mathbb{H}; |x| \leq \frac{1}{2}, |\tau| \geq 1, y \leq T\right\}.$$

Following [7], we define the regularized inner product $\langle f, g\rangle^{\mathrm{reg}}$ of forms f, g which transform like modular forms of weight k, but may grow at the cusps. To be more precise, we let $\langle f, g\rangle^{\mathrm{reg}}$ be the constant term in the Laurent expansion at $s = 0$ of the meromorphic continuation in s of

$$\lim_{T\to\infty} \int_{\mathcal{F}_T} g(\tau)\overline{f(\tau)}y^{k-s} d\mu,$$

if it exists. Note that if $f \in M_k$ and g has vanishing constant term, then we have

$$\langle f, g\rangle^{\mathrm{reg}} = \lim_{T\to\infty} \int_{\mathcal{F}_T} g(\tau)\overline{f(\tau)}y^k d\mu. \tag{1.4}$$

The following theorem may be easily verified.

Theorem 2.8. *The integral* (1.3) *is absolutely convergent and the following conditions hold:*

(i) $\langle f, g\rangle = \overline{\langle g, f\rangle}$*;*

(ii) $\langle f, g\rangle$ *is* $\mathbb{C}$*-linear in* f*;*

(iii) $\langle f, f\rangle \geq 0$ *and moreover* $\langle f, f\rangle = 0 \Leftrightarrow f \equiv 0$.

In particular, integrating against Poincaré series yields the Fourier coefficients of cusp forms.

Theorem 2.9 (Petersson coefficient formula). *For* $f(\tau) = \sum_{m\geq 1} a(m)e^{2\pi i m\tau} \in S_k$*, we have*

$$\langle f, P_{k,n}\rangle = \begin{cases} 0 & \text{for } n = 0, \\ \frac{(k-2)!}{(4\pi n)^{k-1}}a(n) & \text{for } n \geq 1. \end{cases}$$

In particular, $E_k \perp S_k$ *with respect to the Petersson scalar product.*

Proof. We only argue formally and ignore questions of convergence. We have

$$\begin{aligned}\langle f, P_{k,n}\rangle &= \int_{\mathcal{F}} f(\tau) \sum_{M\in\Gamma_\infty\backslash \mathrm{SL}_2(\mathbb{Z})} \overline{e^{2\pi i n\tau}|_k M}\, y^k d\mu \\ &= \sum_{M\in\Gamma_\infty\backslash \mathrm{SL}_2(\mathbb{Z})} \int_{\mathcal{F}} f\big|_k M(\tau)\overline{e^{2\pi i n\tau}|_k M}\, y^k d\mu \\ &= \int_{\bigcup_{M\in\Gamma_\infty\backslash \mathrm{SL}_2(\mathbb{Z})} M\mathcal{F}} f(\tau)\overline{e^{2\pi i n\tau}}\, y^k d\mu. \qquad (1.5)\end{aligned}$$

Note that $\bigcup_{M\in\Gamma_\infty\backslash \mathrm{SL}_2(\mathbb{Z})} M\mathcal{F}$ is a fundamental domain for the action of Γ_∞ on $\mathbb{H}$ and that $f(\tau)\overline{e^{2\pi i n\tau}}y^k$ is invariant under the action of Γ_∞. Thus we may change $\bigcup_{M\in\Gamma_\infty\backslash\Gamma} M\mathcal{F}$ into

$$\{\tau = x+iy; 0\le x\le 1, y>0\}.$$

This gives that (1.5) equals

$$\begin{aligned}&\int_0^\infty\int_0^1 f(\tau)e^{-2\pi i n\overline{\tau}}y^{k-2}dxdy \\ &\quad= \sum_{m\ge 1} a(m)\int_0^\infty e^{-2\pi(n+m)y}y^{k-2}dy\int_0^1 e^{2\pi i(m-n)x}dx.\end{aligned}$$

Now the claim follows by using that

$$\begin{aligned}\int_0^1 e^{2\pi i\ell x}dx &= \begin{cases} 1 & \text{if } \ell=0,\\ 0 & \text{otherwise,}\end{cases}\\ \int_0^\infty e^{-t}t^{k-2}dt &= (k-2)!. \end{aligned}$$ □

To show that the Poincaré series constitute a basis of S_k, we first require the Valence formula (see [31] for a proof).

Theorem 2.10. *If $f\not\equiv 0$ is a meromorphic modular form of weight $k\in\mathbb{Z}$, we have with $\rho := e^{\frac{2\pi i}{3}}$*

$$\mathrm{ord}(f;\infty)+\frac{1}{2}\mathrm{ord}(f;i)+\frac{1}{3}\mathrm{ord}(f;\rho)+\sum_{\substack{z\in\Gamma\backslash\mathbb{H}\\ z\not\equiv i,\rho \pmod{\Gamma}}}\mathrm{ord}(f;z)=\frac{k}{12},$$

where $\mathrm{ord}(f;\infty)=n_0$ *if* $f(\tau)=\sum_{n=n_0}^\infty a(n)q^n$ *with* $a(n_0)\neq 0$.

One can conclude from Theorem 2.10 the following dimension formulas, the proof of which is omitted here.

Corollary 2.11. *For $k \in \mathbb{N}_0$,*

$$\dim M_{2k} = \begin{cases} \left\lfloor \frac{k}{6} \right\rfloor + 1 & \text{if } k \not\equiv 1 \pmod 6, \\ \left\lfloor \frac{k}{6} \right\rfloor & \text{if } k \equiv 1 \pmod 6, \end{cases}$$

$$\dim S_k = \begin{cases} \left\lfloor \frac{k}{6} \right\rfloor & \text{if } k \not\equiv 1 \pmod 6, \\ \left\lfloor \frac{k}{6} \right\rfloor - 1 & \text{if } k \equiv 1 \pmod 6, k \geq 7, \\ 0 & \text{if } k = 1. \end{cases}$$

This easily gives the basis property of the Poincaré series.

Corollary 2.12 (Completeness theorem). *Let $k \geq 4$ even and $d_k := \dim(S_k)$. Then a basis of S_k is given by*

$$\left\{P_{k,n}; n = 1, \ldots, d_k\right\}.$$

In particular, M_k has a basis consisting of Eisenstein series and Poincaré series.

Proof. Set $S := \operatorname{span}\left\{P_{k,1}, \ldots, P_{k,d_k}\right\} \subset S_k$ and let $f \in S_k$ be such that $f \perp S$ with respect to the Petersson inner product. Then f has a Fourier expansion of the form $f(\tau) = \sum_{m \geq d_k+1} a(m) e^{2\pi i m \tau}$. From Corollary 2.11 we know the precise values of d_k, yielding a contradiction to Theorem 2.10. To be more precise, we have for $k \not\equiv 2 \pmod{12}$

$$d_k + 1 = \left\lfloor \frac{k}{12} \right\rfloor + 1 > \frac{k}{12} = \operatorname{ord}(f; \infty) + \sum_{z \in \Gamma \backslash \mathbb{H}} \operatorname{ord}(f; z) \geq d_k + 1.$$

For $k \equiv 2 \pmod{12}$, we get

$$\frac{k}{12} = d_k + 1 + \frac{1}{6}.$$

Thus $\operatorname{ord}(f; \infty) = d_k + 1$ and

$$\sum_{z \in \Gamma \backslash \mathbb{H}} \operatorname{ord}(f; z) = \frac{1}{6},$$

which is impossible. □

We next compute the Fourier coefficients $a_n(m)$ of the Poincaré series $P_{k,n}$. For this, we require some special functions. Define the *Kloosterman sums* by

$$S(m, n; c) := \sum_{a \pmod{c}^\star} e^{2\pi i \frac{am + \overline{a}n}{c}},$$

where the sum runs over all $a \pmod{c}$ that are coprime to c and $\overline{a}$ denotes the multiplicative inverse of $a \pmod{c}$. Moreover, we let J_r be the J-Bessel function of order r, defined by

$$J_r(x) := \sum_{\ell \geq 0} \frac{(-1)^\ell}{\ell! \Gamma(\ell + 1 + r)} \left(\frac{x}{2}\right)^{r+2\ell},$$

where $\Gamma(x)$ denotes the usual gamma-function.

Theorem 2.13. *We have for $n \in \mathbb{N}$*

$$a_n(m) = \left(\frac{m}{n}\right)^{\frac{k-1}{2}} \left(\delta_{m,n} + 2\pi i^{-k} \sum_{c \geq 1} c^{-1} S(n, m; c) J_{k-1}\left(\frac{4\pi\sqrt{mn}}{c}\right)\right), \quad (1.6)$$

where

$$\delta_{m,n} := \begin{cases} 1 & \text{if } m = n, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. We again use that a set of representatives of $\Gamma_\infty \backslash \mathrm{SL}_2(\mathbb{Z})$ is given by

$$\left\{ \begin{pmatrix} \star & \star \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}); (c, d) = 1 \right\}.$$

The contribution for $c = 0$ is easily seen to give the first summand in (1.6). For $c \neq 0$ we use the identity

$$\frac{a\tau + b}{c\tau + d} = \frac{a}{c} - \frac{1}{c^2\left(\tau + \frac{d}{c}\right)}$$

and change $d \mapsto d + mc$, where d runs $\pmod{c}^\star$ and $m \in \mathbb{Z}$. This gives

$$P_{k,n}(\tau) = e^{2\pi i n\tau} + 2\sum_{c \geq 1} c^{-k} \sum_{d \pmod{c}^\star} e^{\frac{2\pi i n a}{c}} \mathcal{F}\left(\tau + \frac{d}{c}\right),$$

where a is defined by $ad \equiv 1 \pmod{c}$ and

$$\mathcal{F}(\tau) := \sum_{m \in \mathbb{Z}} e^{-\frac{2\pi i n}{c^2(\tau+m)}} (\tau + m)^{-k}.$$

Now the classical Poisson summation formula yields

$$\mathcal{F}(\tau) = \sum_{m \in \mathbb{Z}} a(m) e^{2\pi i m\tau},$$

where

$$a(m) = \int_{\mathrm{Im}(\tau) = \mathcal{C}} \tau^{-k} e^{-\frac{2\pi i n}{c^2\tau} - 2\pi i m\tau} d\tau$$

with $\mathcal{C} > 0$ arbitrary. For $m \leq 0$ we can deform the path of integration up to infinity yielding that $a(m) = 0$ in this case. For $m > 0$ we make the substitution $\tau = ic^{-1}(n/m)^{\frac{1}{2}}w$ to get

$$a(m) = i^{-k-1}c^{k-1}\left(\frac{m}{n}\right)^{\frac{k}{2}-\frac{1}{2}} \int_{\mathcal{C}-i\infty}^{\mathcal{C}+i\infty} w^{-k} e^{\frac{2\pi}{c}\sqrt{mn}(w-w^{-1})} dw.$$

The claim follows using the fact that for $\mu, \kappa > 0$ the functions

$$t \longmapsto (t/\kappa)^{\frac{\mu-1}{2}} J_{\mu-1}\left(2\sqrt{\kappa t}\right), \quad (t > 0),$$

and

$$w \longmapsto w^{-\mu}e^{-\frac{\kappa}{w}}, \quad (\mathrm{Re}(w) > 0),$$

are inverses of each other with respect to the usual Laplace transform (8.412.2 of [23]). □

3 Weakly holomorphic modular forms

We next turn to *weakly holomorphic* modular forms, which are still holomorphic on $\mathbb{H}$, but admit poles at the "cusps". The Fourier coefficients of such forms grow much faster than those of holomorphic forms. Let us in particular describe this in the situation of the partition function.

Recall that a *partition* of a positive integer n is a nondecreasing sequence of positive integers (the *parts* of the partition) whose sum is n. Let $p(n)$ denote the number of partitions of n. For example, the partitions of 4 are

$$4, \qquad 3+1, \qquad 2+2, \qquad 2+1+1, \qquad 1+1+1+1,$$

so that $p(4) = 5$. The partition function is very rapidly increasing. For example,

$$\begin{aligned} p(3) &= 3, \\ p(4) &= 5, \\ p(10) &= 42, \\ p(20) &= 627, \\ p(100) &= 190569292. \end{aligned}$$

A key observation by Euler is the product identity

$$P(q) := 1 + \sum_{n\geq 1} p(n)q^n = \prod_{n\geq 1} \frac{1}{1-q^n}.$$

One can show that for $|q| < 1$, the function P is holomorphic. Using Euler's identity one can embed the partition function into the modular world using the Dedekind η-function ($q = e^{2\pi i \tau}$)

$$\eta(\tau) := q^{\frac{1}{24}} \prod_{n\geq 1} (1 - q^n) \,.$$

This function is a modular form of weight $1/2$. To be more precise, we have the following transformation laws (see e.g. [31]).

Theorem 3.1. *We have*

$$\eta(\tau + 1) = e^{\frac{\pi i}{12}} \eta(\tau),$$
$$\eta\left(-\frac{1}{\tau}\right) = \sqrt{-i\tau}\eta(\tau).$$

Theorem 3.1 in particular gives the asymptotic behavior of η as $\tau \to 0$. To be more precise, it implies that

$$P(q) \sim \sqrt{-i\tau} e^{\frac{\pi i}{12\tau}} \qquad (\tau \to 0).$$

We moreover note that the coefficients $p(n)$ are easily seen to be positive and monotonic. From this, we may conclude the growth behavior of $p(n)$ using a Tauberian Theorem due to Ingham [27].

Theorem 3.2. *Assume that $f(\tau) := q^{n_0} \sum_{n\geq 0} a(n) q^n$ is a holomorphic function on $\mathbb{H}$, satisfying the following conditions:*

(i) *For all $n \in \mathbb{N}_0$, we have*

$$0 < a(n) \leq a(n + 1).$$

(ii) *There exist $c \in \mathbb{C}$, $d \in \mathbb{R}$, and $N > 0$ such that*

$$f(\tau) \sim c(-i\tau)^{-d} e^{\frac{2\pi i N}{\tau}} \qquad (\tau \to 0).$$

Then

$$a(n) \sim \frac{c}{\sqrt{2} N^{\frac{1}{2}\left(d - \frac{1}{2}\right)}} n^{\frac{1}{2}\left(d - \frac{3}{2}\right)} e^{4\pi\sqrt{Nn}} \qquad (n \to \infty).$$

From this we immediately conclude the growth behavior of the partition function.

Theorem 3.3. *We have*

$$p(n) \sim \frac{1}{4n\sqrt{3}} \cdot e^{\pi\sqrt{\frac{2n}{3}}} \qquad (n \to \infty). \tag{1.7}$$

We note that the partition function also has the q-hypergeometric series representation

$$P(q) = \sum_{n\geq 0} \frac{q^{n^2}}{(q;q)_n^2}, \tag{1.8}$$

where $(a;q)_n := \prod_{j=0}^{n-1}(1-aq^j)$. Showing modularity by just using this representation is still an open problem [3].

Rademacher [36], building on work of Hardy and Ramanujan [22], used the Circle Method to obtain an exact formula for $p(n)$. To state his result, we let

$$I_s(x) := i^{-s} J_s(ix) = \sum_{m\geq 0} \frac{1}{m!\Gamma(m+\alpha+1)} \left(\frac{x}{2}\right)^{2m+\alpha} \tag{1.9}$$

be the *I-Bessel function of order s*. Moreover, with $\chi_{12}(x) := \left(\frac{12}{x}\right)$, we define the *Kloosterman sum*

$$A_k(n) := \frac{\sqrt{k}}{4\sqrt{3}} \sum_{\substack{x \pmod{24k} \\ x^2\equiv 1-24n \pmod{24k}}} \chi_{12}(x) e^{\frac{2\pi i x}{12k}}.$$

Note that for $k>0$, $(h,k)=1$, and $\mathrm{Re}(z)>0$ we may rewrite the transformation law of the partition generating function as

$$P\left(\exp\left(\frac{2\pi i}{k}(h+iz)\right)\right) = \omega_{h,k}\sqrt{z}e^{\frac{\pi}{12k}\left(z^{-1}-z\right)} P\left(\exp\left(\frac{2\pi i}{k}\left(h'+\frac{i}{z}\right)\right)\right), \tag{1.10}$$

where $hh' \equiv -1 \pmod{k}$. Here

$$\omega_{h,k} := \exp\left(\pi i s\left(h,k\right)\right), \tag{1.11}$$

with

$$s\left(h,k\right) := \sum_{\mu \pmod{k}} \left(\left(\frac{\mu}{k}\right)\right)\left(\left(\frac{h\mu}{k}\right)\right)$$

and

$$((x)) := \begin{cases} x - \lfloor x \rfloor - \frac{1}{2} & \text{if } x \in \mathbb{R}\backslash\mathbb{Z}, \\ 0 & \text{if } x \in \mathbb{Z}. \end{cases}$$

Sometimes it is also useful to rewrite $\omega_{h,k}$ as [33]

$$\omega_{h,k} = \begin{cases} \left(\frac{-k}{h}\right) e^{-\pi i\left(\frac{1}{4}(2-hk-h)+\frac{1}{12}\left(k-k^{-1}\right)\left(2h-h'+h^2h'\right)\right)} & \text{if } h \text{ is odd}, \\ \left(\frac{-h}{k}\right) e^{-\pi i\left(\frac{1}{4}(k-1)+\frac{1}{12}\left(k-k^{-1}\right)\left(2h-h'+h^2h'\right)\right)} & \text{if } k \text{ is odd}. \end{cases}$$

Here $\left(\frac{a}{b}\right)$ denotes the Jacobi symbol.

Note that we may also write

$$A_k(n) = \sum_{h \pmod{k^\star}} \omega_{h,k} e^{-\frac{2\pi i n h}{k}}. \tag{1.12}$$

Theorem 3.4. *For $n \geq 1$, we have*

$$p(n) = \frac{2\pi}{(24n-1)^{\frac{3}{4}}} \sum_{k\geq 1} \frac{A_k(n)}{k} I_{\frac{3}{2}}\left(\frac{\pi\sqrt{24n-1}}{6k}\right).$$

We note that this is a very astonishing identity expressing the integer $p(n)$ as an infinite sum of transcendental numbers. Recently Bruinier and Ono [16] found a formula expressing $p(n)$ as a finite sum of algebraic numbers.

Proof. Here we only give some details of the proof, for more see [4]. By Cauchy's Theorem,

$$p(n)=\frac{1}{2\pi i}\int_{\mathcal{C}}\frac{P(q)}{q^{n+1}}dq,$$

where $\mathcal{C}$ is any path inside the unit circle surrounding 0 counterclockwise. We may choose for $\mathcal{C}$ the circle centered at 0 with radius $\rho=\exp(-\frac{2\pi}{N^2})$, with $N>0$ fixed (later we let $N\to\infty$). Then

$$p(n)=\rho^{-n}\int_0^1 P\left(\rho\exp\left(2\pi i t\right)\right)\exp\left(-2\pi i n t\right)dt.$$

Define

$$\vartheta'_{h,k}:=\frac{1}{k\left(k_1+k\right)}\qquad\text{and}\qquad\vartheta''_{h,k}:=\frac{1}{k\left(k_2+k\right)},$$

where $\frac{h_1}{k_1}<\frac{h}{k}<\frac{h_2}{k_2}$ are adjacent Farey fractions in the Farey sequence of order N (for example, see p. 72 of [4]). From the theory of Farey fractions it is known that $(j=1,2)$

$$\frac{1}{k+k_j}\leq\frac{1}{N+1}.$$

We decompose the path of integration in paths along the Farey arcs $-\vartheta'_{h,k}\leq\phi\leq\vartheta''_{h,k}$, where $\phi=t-\frac{h}{k}$. Setting $z=k(N^{-2}-i\phi)$ then yields

$$p(n)=\exp\left(\frac{2\pi n}{N^2}\right)\sum_{\substack{1\leq k\leq N\\0\leq h<k\\(h,k)=1}}\exp\left(-\frac{2\pi i n h}{k}\right)$$
$$\times\int_{-\vartheta'_{h,k}}^{\vartheta''_{h,k}}P\left(\exp\left(\frac{2\pi i}{k}(h+iz)\right)\right)\exp\left(-2\pi i n\phi\right)d\phi.$$

For $N\to\infty$ we have that $z\to0$, i.e., $\exp(\frac{2\pi i}{k}(h+iz))\to\exp\left(\frac{2\pi i h}{k}\right)$. We thus need to know the behavior of $P(q)$ as $q\to\exp\left(\frac{2\pi i h}{k}\right)$. To find it, we apply (1.10) to obtain

$$p(n)=\exp\left(\frac{2\pi n}{N^2}\right)\sum_{\substack{1\leq k\leq N\\0\leq h<k\\(h,k)=1}}\exp\left(-\frac{2\pi i n h}{k}\right)\omega_{h,k}$$
$$\times\int_{-\vartheta'_{h,k}}^{\vartheta''_{h,k}}z^{\frac{1}{2}}\exp\left(\frac{\pi}{12k}(z^{-1}-z)\right)P\left(\exp\left(\frac{2\pi i}{k}\left(h'+\frac{i}{z}\right)\right)\right)\exp\left(-2\pi i n\phi\right)d\phi.$$

Now $\exp(\frac{2\pi i}{k}(h' + \frac{i}{z})) \to 0$ as $z \to 0^+$. Thus all the terms in $P(q) - 1$ are "small". We therefore write

$$p(n) = \sum\nolimits_1 + \sum\nolimits_2,$$

with

$$\sum\nolimits_1 := \exp\left(\frac{2\pi n}{N^2}\right) \sum_{\substack{1\le k\le N\\ 0\le h<k\\ (h,k)=1}} \exp\left(-\frac{2\pi i n h}{k}\right) \omega_{h,k}$$

$$\times \int_{-\vartheta'_{h,k}}^{\vartheta''_{h,k}} z^{\frac{1}{2}} \exp\left(\frac{\pi}{12k}\left(z^{-1} - z\right)\right) \exp\left(-2\pi i n\phi\right) d\phi,$$

$$\sum\nolimits_2 := \exp\left(\frac{2\pi n}{N^2}\right) \sum_{\substack{1\le k\le N\\ 0\le h<k\\ (h,k)=1}} \exp\left(-\frac{2\pi i n h}{k}\right) \omega_{h,k} \int_{-\vartheta'_{h,k}}^{\vartheta''_{h,k}} z^{\frac{1}{2}} \exp\left(\frac{\pi\left(z^{-1} - z\right)}{12k}\right)$$

$$\times \left(P\left(\exp\left(\frac{2\pi i}{k}\left(h' + \frac{i}{z}\right)\right)\right) - 1\right) \exp\left(-2\pi i n\phi\right) d\phi.$$

Here we only focus on the main term $\sum_1$ and only note that $\sum_2$ contributes to the error terms, giving

$$\sum\nolimits_2 = O\left(N^{-\frac{1}{2}} \exp\left(\frac{2\pi n}{N^2}\right)\right) \to 0$$

for $N \to \infty$. The proof is given in [4].

Turning to the main term, setting $w := N^{-2} - i\phi$ yields

$$\sum\nolimits_1 = \exp\left(\frac{2\pi n}{N^2}\right) \sum_{\substack{1\le k\le N\\ 0\le h<k\\ (h,k)=1}} \exp\left(-\frac{2\pi i n h}{k}\right) \omega_{h,k} I_{h,k},$$

with

$$I_{h,k} := -i k^{\frac{1}{2}} \exp\left(-2\pi n N^{-2}\right) \int_{N^{-2}-i\vartheta''_{h,k}}^{N^{-2}+i\vartheta'_{h,k}} g(w) dw.$$

Here

$$g(w) := w^{\frac{1}{2}} \exp\left(2\pi\left(n - \frac{1}{24}\right) w + \frac{\pi}{12k^2 w}\right).$$

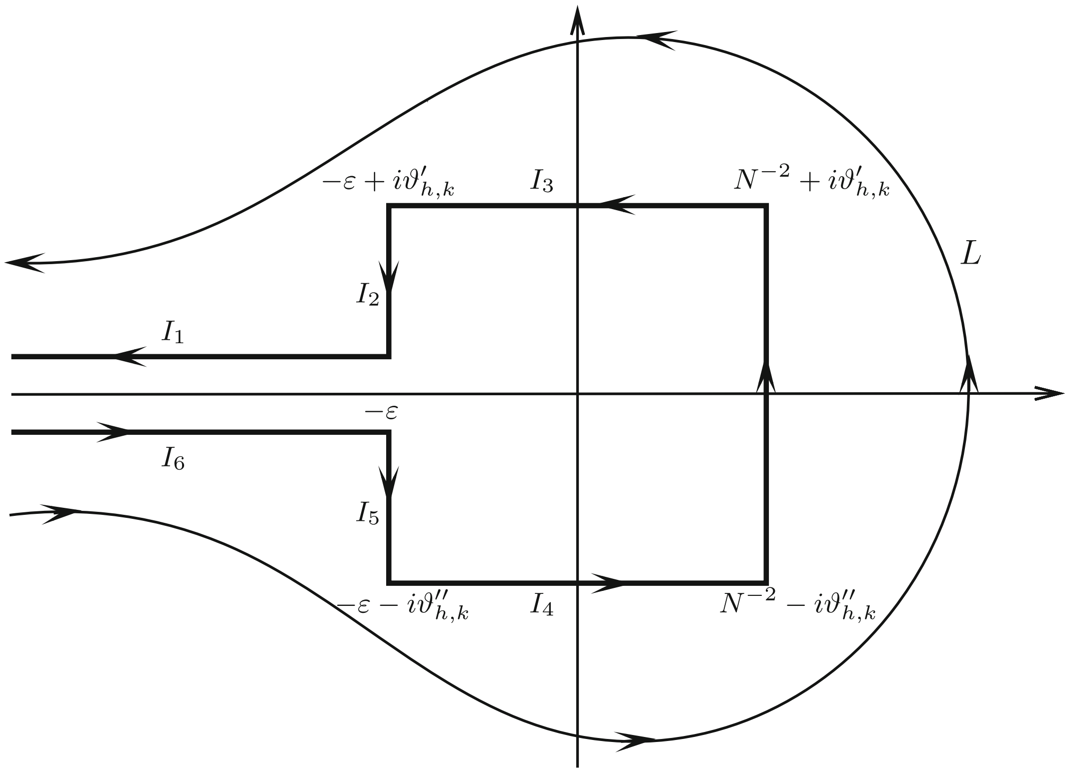
Using the Residue Theorem, we can write

$$\exp\left(\frac{2\pi n}{N^2}\right) I_{h,k} = -k^{\frac{1}{2}} i \left(\mathcal{L}_k - \mathcal{I}_1 - \mathcal{I}_2 - \mathcal{I}_3 - \mathcal{I}_4 - \mathcal{I}_5 - \mathcal{I}_6\right),$$

with

$$\mathcal{L}_k := \int_L z^{\frac{1}{2}} \exp\left(2\pi\left(n-\frac{1}{24}\right)z+\frac{\pi}{12k^2z}\right)dz,$$
$$\mathcal{I}_j := \int_{I_j} z^{\frac{1}{2}} \exp\left(2\pi\left(n-\frac{1}{24}\right)z+\frac{\pi}{12k^2z}\right)dz,$$

where L and I_j denote the paths of integration in the picture


$-\varepsilon + i\vartheta'_{h,k}$
I_3
$N^{-2} + i\vartheta'_{h,k}$
L
I_2
I_1
$-\varepsilon$
I_6
I_5
$-\varepsilon - i\vartheta''_{h,k}$
I_4
$N^{-2} - i\vartheta''_{h,k}$

We note without a proof that $\mathcal{I}_2$, $\mathcal{I}_3$, $\mathcal{I}_4$, and $\mathcal{I}_5$ contribute to the error and vanish as $N \to \infty$. Moreover, as $\varepsilon \to 0$

$$\mathcal{I}_1 + \mathcal{I}_6 = -2i \int_0^\infty t^{\frac{1}{2}} \exp\left(-2\pi\left(n-\frac{1}{24}\right)t-\frac{\pi}{12k^2t}\right)dt =: -2i\mathcal{L}_k^\star$$

This gives that

$$p(n) = \sum_{k\geq 1} A_k(n)\psi_k(n),$$

where $A_k(n)$ is defined as in (1.12) and

$$\psi_k(n) := -i\sqrt{k}\mathcal{L}_k + 2\sqrt{k}\mathcal{L}_k^\star.$$

To finish the proof, we have to show that

$$\psi_k(n) = \frac{2\pi}{(24n-1)^{\frac{3}{4}}} \frac{1}{k} I_{\frac{3}{2}}\left(\frac{\pi\sqrt{24n-1}}{6k}\right).$$

Inserting the power series expansion for the exponential function, interchanging summation and integration, and then making a change of variables gives that

$$-i\mathcal{L}_k = 2\pi \sum_{s\geq 0} \frac{\left(\frac{\pi}{12k^2}\right)^s}{s!} \left(2\pi\left(n-\frac{1}{24}\right)\right)^{s-\frac{3}{2}} \frac{1}{2\pi i}\int_L e^z z^{-s+\frac{1}{2}} dz,$$

where the loop L is as in the above picture.

Using the Hankel loop integral formula

$$\frac{1}{\Gamma\left(s-\frac{1}{2}\right)} = \frac{1}{2\pi i}\int_L e^z z^{-s+\frac{1}{2}} dz$$

then yields that

$$-i\mathcal{L}_k = \frac{1}{\sqrt{2\pi}} \frac{1}{\left(n-\frac{1}{24}\right)^{\frac{3}{2}}} \sum_{s\geq 0} \frac{\left(\frac{\pi^2\left(n-\frac{1}{24}\right)}{6k^2}\right)^s}{s!\Gamma\left(s-\frac{1}{2}\right)}.$$

Treating $\mathcal{L}_k^\star$ similarly, the claim follows using the series representation of the Bessel function (1.9). □

Corollary 3.5. *The asymptotic estimate* (1.7) *is true.*

Proof. The claim follows immediately from Theorem 3.4 using that

$$I_\ell(x) \sim \frac{e^x}{\sqrt{2\pi x}} \qquad (x \to \infty).$$ □

4 Mock modular forms

Next we aim to study coefficients of *mock modular forms*, which are holomorphic parts of *harmonic* (*weak*) *Maass forms*. These are a generalization of modular forms which satisfy a transformation law like (1.1) and (weak) growth conditions at the cusps, but, instead of being meromorphic, they are annihilated by the weight-k *hyperbolic Laplacian* ($\tau = x + iy$)

$$\Delta_k := -y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + iky\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right).$$

We let H_k denote the space of harmonic Maass forms of weight k. A theory of harmonic Maass forms was built by Bruinier and Funke [15]. Previously, Niebur [34, 35] and Hejhal [24] constructed certain non-holomorphic Poincaré series which turn out to be examples of harmonic weak Maass forms. Since functions that are holomorphic on $\mathbb{H}$ are annihilated by Δ_k, weakly holomorphic modular forms are harmonic Maass forms. The simplest (non-weakly holomorphic) harmonic weak Maass form is given by the non-holomorphic Eisenstein series of weight 2. To be more precise, define

$$E_2(\tau) := 1 - 24 \sum_{n \geq 1} \sigma_1(n) q^n.$$

Then E_2 is translation invariant, but introduces an error term under modular inversion:

$$E_2\left(-\frac{1}{\tau}\right) = \tau^2 E_2(\tau) + \frac{6\tau}{\pi i}.$$

This gives that

$$\widehat{E}_2(\tau) := E_2(\tau) - \frac{3}{\pi y}$$

is a harmonic weak Maass form of weight 2.

A further example of a mock modular form of weight $\frac{3}{2}$ is given by the generating function of Hurwitz class numbers [25]. To be more precise, the function

$$\widehat{h}(\tau) := \sum_{\substack{n \geq 0 \\ n \equiv 0,3 \pmod 4}} H(n) q^n + \frac{i}{8\sqrt{2}\pi} \int_{-\overline{\tau}}^{i\infty} \frac{\Theta(w)}{(-i\,(\tau + w))^{\frac{3}{2}}} dw \tag{1.13}$$

is a harmonic Maass forms of weight $\frac{3}{2}$ on $\Gamma_0(4)$. Here $H(n)$ is the nth *Hurwitz class number*, i.e., the number of equivalence classes of quadratic forms of discriminant $-n$, where each class C is counted with multiplicity $1/\mathrm{Aut}(C)$ and $\Theta(w) := \sum_{n \in \mathbb{Z}} e^{2\pi i n^2 w}$ is the usual weight-$\frac{1}{2}$ theta function. We note that these class numbers may also be related to so-called over partitions [9].

In this paper, we are in particular interested in those harmonic Maass forms $\mathcal{F}$ for which there exist a polynomial $\mathcal{P}_{\mathcal{F}}$ and $\ell > 0$ such that as $y \to \infty$

$$\mathcal{F}(\tau) - \mathcal{P}_{\mathcal{F}}\left(q^{-1}\right) = O\left(e^{-\ell y}\right).$$

The function $\mathcal{P}_{\mathcal{F}}(q^{-1})$ is called the *principal part* of $\mathcal{F}$. We denote the associated space of Maass forms of weight k by H_k^*.

Using the differential equation satisfied by harmonic Maass forms, it is not hard to see that forms in H_k^* naturally split into a holomorphic and a non-holomorphic part:

$$\mathcal{F}(\tau) = \mathcal{F}^+(\tau) + \mathcal{F}^-(\tau),$$

where

$$\mathcal{F}^{+}(\tau) := \sum_{n \gg -\infty} a^{+}(n) q^{n},$$

$$\mathcal{F}^{-}(\tau) := \sum_{n<0} a^{-}(n) \Gamma(1-k; 4\pi|n|y) q^{n}.$$

Here

$$\Gamma(\alpha; x) := \int_{x}^{\infty} e^{-t} t^{\alpha-1} dt$$

is the *incomplete gamma function*. The function $\mathcal{F}^{+}$ is called a *mock modular form* and its coefficients often encode interesting arithmetic information. Each mock modular form has a hidden companion, its *shadow*, which is necessary to fully understand the mock modular form. The shadow, which is a classical cusp form of weight $2-k$, may be obtained from the associated harmonic weak Maass form of weight k by applying the differential operator $\xi_k := 2iy^k \overline{\frac{\partial}{\partial \overline{\tau}}}$. A direct calculation gives

$$\xi_k(\mathcal{F}) = -(4\pi)^{1-k} \sum_{n \geq 1} \overline{a^{-}(-n)} n^{1-k} q^{n}.$$

A space "orthogonal" to H_k^* can be charaterized in terms of the holomorphic differential operator

$$D^{k-1} := \left(\frac{1}{2\pi i} \frac{\partial}{\partial z}\right)^{k-1}.$$

Bruinier, Ono, and Rhoades [17] showed the following

Theorem 4.1. *If $k \in \mathbb{N}, k \geq 2$, then the image of the map*

$$D^{k-1} : H_{2-k} \to M_k^!$$

consists of those $h \in M_k^!$ which are orthogonal to cusp forms with respect to the regularized inner product (1.4)*, and which also have constant term* 0.

Further examples of mock modular forms are given by the so-called *mock theta functions*, a collection of 22 q-series including

$$f(q) := \sum_{n \geq 0} \frac{q^{n^2}}{(-q;q)_n^2} = \sum_{n \geq 0} \alpha(n) q^{n},$$

which were defined by the visionary Ramanujan in his last letter to Hardy [37]. Note that this function agrees with the representation (1.8) of $P(q)$ as a q-hypergeometric function up to a simple change of sign. Ramanujan's last letter to Hardy includes the claim that

$$\alpha(n) \sim \frac{(-1)^{n+1}}{2\sqrt{n}} e^{\pi\sqrt{\frac{n}{6}}} \qquad (\text{as } n \to \infty).$$

As typical for his writing, Ramanujan left no proof of this claim. Dragonette, a Ph.D. student of Rademacher, then solved this claim in her thesis [19]. Andrews [1] improved upon her work and together they made the following conjecture.

Conjecture (Andrews, Dragonette). *For $n \in \mathbb{N}$, we have*

$$\alpha(n) = \frac{\pi}{(24n-1)^{\frac{1}{4}}} \sum_{k\geq 1} \frac{(-1)^{\left\lfloor \frac{k+1}{2} \right\rfloor} A_{2k}\left(n - \frac{k\left(1+(-1)^k\right)}{4}\right)}{k} I_{\frac{1}{2}}\left(\frac{\pi\sqrt{24n-1}}{12n}\right).$$

Note that not even convergence of this exact formula is obvious. In joint work with K. Ono, we proved this conjecture [12].

Theorem 4.2. *The Andrews-Dragonette Conjecture is true.*

We note that S. Garthwaite [21] showed in her thesis a similar result for the coefficients of Ramanujan's mock theta function $\omega(q)$ defined in (1.14). Later we [13] proved exact formulas for coefficients of a generic harmonic weak Maass form.

The coefficients $\alpha(n)$ also have some combinatorial interpretations. To describe this, recall that the *rank* of a partition is its largest part minus the number of its parts. Dyson [20] introduced this statistic to explain the famous Ramanujan congruences

$$\begin{aligned} p(5n+4) &\equiv 0 \pmod{5}, \\ p(7n+5) &\equiv 0 \pmod{7}, \\ p(11n+6) &\equiv 0 \pmod{11}. \end{aligned}$$

More precisely, Dyson conjectured that the partitions of $5n+4$ (resp. $7n+5$) form 5 (resp. 7) groups of equal size when sorted by their ranks modulo 5 (resp. 7). As an example, the following table gives the ranks for the partitions of 4.

partition	rank	rank (mod 5)
4	$4-1=3$	3
$3+1$	$3-2=1$	1
$2+2$	$2-2=0$	0
$2+1+1$	$2-3=-1$	4
$1+1+1+1$	$1-4=-3$	2

Since a direct calculation confirms that such a splitting in residue classes cannot hold modulo 11, Dyson postulated the existence of another statistic, which should be called the *crank* and which should explain all of the Ramanujan congruences. Dyson's rank conjecture was solved by Atkin and Swinnerton-Dyer [6]; the crank was found by Andrews and Garvan [5]. To study ranks it is natural to consider a generating function. Denoting by $N(m,n)$ the number of partitions of n of rank m, it is well-known that

$$1+\sum_{n\geq 1}\sum_{m\in\mathbb{Z}} N\left(m,n\right) z^m q^n = 1+\sum_{n\geq 1}\frac{q^{n^2}}{(zq;q)_n\left(z^{-1}q;q\right)_n}.$$

In particular, letting $z=-1$, we obtain

$$1+\sum_{n\geq 1}\left(N_{\mathrm{e}}(n)-N_{\mathrm{o}}(n)\right)q^n = 1+\sum_{n\geq 1}\frac{q^{n^2}}{(-q;q)_n^2} = f(q),$$

where $N_{\mathrm{e}}(n)$ (resp. $N_{\mathrm{o}}(n)$) denotes the number of partitions of n of even (resp. odd) rank. This yields the combinatorial interpretation of the coefficients $\alpha(n)$ of the mock theta function $f(q)$.

Proof of Theorem 4.2. The idea in [12] is to realize $f(q)$ as the holomorphic part of a Maass-Poincaré series. Such Poincaré series generate the space of harmonic Maass forms and have the form

$$\sum_{M=\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)\in\Gamma_\infty\backslash\Gamma}\left(c\tau+d\right)^{-k}\phi\left(\frac{a\tau+b}{c\tau+d}\right),$$

where Γ is an appropriate subgroup of $\mathrm{SL}_2(\mathbb{Z})$ and ϕ is a translation-invariant function that satisfies the differential equation of a Maass form. Fourier coefficients of such Maass-Poincaré series are then easily computed using Poisson summation.

To be more precise, we need to recall work of Zwegers [39] which completes $f(q)$ to a (vector-valued) harmonic weak Maass form. Watson [38] had previously obtained (mock) transformations for $f(q)$ and related functions. To state Zwegers's result we require a further mock theta function

$$\omega(q) := \sum_{n\geq 0}\frac{q^{2n^2+2n}}{(q;q^2)_{n+1}^2} \tag{1.14}$$

to build the vector-valued function

$$F(\tau) := \left(F_0(\tau),F_1(\tau),F_2(\tau)\right)^T = \left(q^{-\frac{1}{24}}f(q), 2q^{\frac{1}{3}}\omega\left(q^{\frac{1}{2}}\right), 2q^{\frac{1}{3}}\omega\left(-q^{\frac{1}{2}}\right)\right)^T.$$

We moreover require its non-holomorphic companion, a certain "period integral",

$$G(\tau) := 2i\sqrt{3}\int_{-\overline{\tau}}^{i\infty}\frac{\left(g_1(z),g_0(z),-g_2(z)\right)^T}{\sqrt{-i(z+\tau)}}dz,$$

where the following cuspidal theta functions of weight $\frac{3}{2}$ are defined as

$$\begin{aligned} g_0(z) &:= \sum_{n\in\mathbb{Z}} (-1)^n \left(n+\frac{1}{3}\right) e^{3\pi i \left(n+\frac{1}{3}\right)^2 z}, \\ g_1(z) &:= -\sum_{n\in\mathbb{Z}} \left(n+\frac{1}{6}\right) e^{3\pi i \left(n+\frac{1}{6}\right)^2 z}, \\ g_2(z) &:= \sum_{n\in\mathbb{Z}} \left(n+\frac{1}{3}\right) e^{3\pi i \left(n+\frac{1}{3}\right)^2 z}. \end{aligned}$$

Then define the completion of F by

$$\widehat{F}(\tau) := F(\tau) - G(\tau).$$

Theorem 4.3. *We have the following transformation laws for the generators of* $\mathrm{SL}_2(\mathbb{Z})$*:*

$$\widehat{F}(\tau+1) = \begin{pmatrix} \zeta_{24}^{-1} & 0 & 0 \\ 0 & 0 & \zeta_3 \\ 0 & \zeta_3 & 0 \end{pmatrix} \widehat{F}(\tau),$$

$$\widehat{F}\left(-\frac{1}{\tau}\right) = \sqrt{-i\tau} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \widehat{F}(\tau),$$

where $\zeta_n := e^{\frac{2\pi i}{n}}$*. Moreover,*

$$\Delta_{\frac{1}{2}}\left(\widehat{F}\right) = 0.$$

From Theorem 4.3 it is then not hard to conclude that

$$\widehat{F}_0(\tau) := F_0(24\tau) - G_0(24\tau)$$

is a harmonic weak Maass form of weight $\frac{1}{2}$ on $\Gamma_0(144)$ with Nebentypus character $\chi_{12}(n) := \left(\frac{12}{n}\right)$. Here $\Gamma_0(N) := \left\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}); c \equiv 0 \pmod{N}\right\}$.

Let us next define the precise Maass-Poincaré series. For this, we require further notation. For matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(2)$ with $c \geq 0$, define the multiplier

$$\chi\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) := \begin{cases} \zeta_{24}^{-b} & \text{if } c = 0, \\ i^{-\frac{1}{2}} (-1)^{\frac{1}{2}(c+ad+1)} e\left(-\frac{a+d}{24c} - \frac{a}{4} + \frac{3dc}{8}\right) \omega_{-d,c}^{-1} & \text{if } c > 0, \end{cases}$$

where $\omega_{h,k}$ was given in (1.11). Here $e(x) := e^{2\pi i x}$. Note that this multiplier is defined to agree with the automorphy factor of $\widehat{F}_0$ when restricted to $\Gamma_0(2)$.

A key function for building the Maass-Poincaré series is the special function ($s \in \mathbb{C}$, $u \in \mathbb{R}\backslash\{0\}$)

$$\mathcal{M}_s(u) := |u|^{-\frac{1}{4}} M_{\frac{1}{4}\operatorname{sgn}(u), s-\frac{1}{2}}\left(|u|\right).$$

Here $M_{\nu,\mu}$ is the standard M-Whittaker function, which satisfies the differential equation

$$\frac{\partial^2}{\partial u^2}\mathcal{F} + \left(-\frac{1}{4} + \frac{\nu}{u} + \frac{\frac{1}{4}-\mu^2}{u^2}\right)\mathcal{F} = 0.$$

Then set

$$\varphi_s(\tau) := \mathcal{M}_s\left(-\frac{\pi y}{6}\right) e\left(-\frac{x}{24}\right).$$

A direct calculation shows that

$$\Delta_{\frac{1}{2}}\left(\varphi_s\right) = \left(s-\frac{1}{4}\right)\left(\frac{3}{4}-s\right)\varphi_s. \tag{1.15}$$

From the function φ_s we then build the Poincaré series

$$P_{\frac{1}{2}}\left(s;\tau\right) := \frac{2}{\sqrt{\pi}} \sum_{M\in\Gamma_\infty\backslash\Gamma_0(2)} \chi(M)^{-1}(c\tau+d)^{-\frac{1}{2}}\varphi_s\left(M\tau\right),$$

where we always choose representations with non-negative lower-left entry. For $\mathrm{Re}(s) > 1$, this series converges absolutely and satisfies for $M = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in \Gamma_0(2)$ the correct transformation law

$$P_{\frac{1}{2}}\left(s;\frac{a\tau+b}{c\tau+d}\right) = \chi(M)(c\tau+d)^{\frac{1}{2}}P_{\frac{1}{2}}\left(s;\tau\right).$$

Specializing to $s = \frac{3}{4}$ then formally yields, by (1.15), a function that is annihilated by $\Delta_{\frac{1}{2}}$. Convergence is however not guaranteed. To overcome this problem, we analytically continue the function via its Fourier expansion.

Proposition 4.4. *We have*

$$P_{\frac{1}{2}}\left(\frac{3}{4};\tau\right) = \left(1-\frac{1}{\sqrt{\pi}}\Gamma\left(\frac{1}{2};\frac{\pi y}{6}\right)\right)q^{-\frac{1}{24}} + \sum_{n\geq 1}\beta(n)q^{n-\frac{1}{24}} + \sum_{n\leq 0}\gamma(n)\Gamma\left(\frac{1}{2};\frac{\pi\left|24n-1\right|y}{6}\right)q^{n-\frac{1}{24}}.$$

Here

$$\beta(n) = \frac{\pi}{(24n-1)^{\frac{1}{4}}}\sum_{k\geq 1}\frac{(-1)^{\lfloor\frac{k+1}{2}\rfloor}A_{2k}\left(n-\frac{k\left(1+(-1)^k\right)}{4}\right)}{k}I_{\frac{1}{2}}\left(\frac{\pi\sqrt{24n-1}}{12k}\right),$$

$$\gamma(n) = \frac{\sqrt{\pi}}{|24n-1|^{\frac{1}{4}}}\sum_{k\geq 1}\frac{(-1)^{\lfloor\frac{k+1}{2}\rfloor}A_{2k}\left(n-\frac{k\left(1+(-1)^k\right)}{4}\right)}{k}J_{\frac{1}{2}}\left(\frac{\pi\sqrt{|24n-1|}}{12k}\right).$$

Proof. Using Poisson summation, the proof proceeds similarly to the proof of Theorem 2.13. In particular we require the following integral evaluations (see page 357 of [24]):

$$\int_{-\infty}^{\infty}\left(\frac{1-iu}{1+iu}\right)^{-\frac{k}{2}} M_{-\frac{k}{2},s-\frac{1}{2}}\left(\frac{4\pi B}{1+u^2}\right)\exp\left(\frac{2\pi i Bu}{1+u^2}+2\pi i Au\right)du$$

$$=\Gamma(2s)\begin{cases}\frac{2\pi\sqrt{\left|\frac{B}{A}\right|}}{\Gamma\left(s-\frac{k}{2}\right)}W_{-\frac{k}{2},s-\frac{1}{2}}\left(4\pi|A|\right)J_{2s-1}\left(4\pi\sqrt{|AB|}\right) & \text{if } A>0,\\ \frac{4\pi^{1+s}}{(2s-1)\Gamma\left(s+\frac{k}{2}\right)\Gamma\left(s-\frac{k}{2}\right)}B^s & \text{if } A=0,\\ \frac{2\pi\sqrt{\left|\frac{B}{A}\right|}}{\Gamma\left(s+\frac{k}{2}\right)}W_{\frac{k}{2},s-\frac{1}{2}}\left(4\pi|A|\right)I_{2s-1}\left(4\pi\sqrt{|AB|}\right) & \text{if } A<0.\end{cases}$$

Here $W_{\nu,\mu}$ is the usual W-Whittaker function. The claim then follows from the special values of Whittaker functions ($y>0$)

$$\mathcal{W}_{\frac{3}{4}}(y)=e^{-\frac{y}{2}},$$
$$\mathcal{W}_{\frac{3}{4}}(-y)=e^{\frac{y}{2}}\Gamma\left(\frac{1}{2};y\right),$$
$$\mathcal{M}_{\frac{3}{4}}(-y)=\frac{1}{2}\left(\sqrt{\pi}-\Gamma\left(\frac{1}{2};y\right)\right)e^{\frac{y}{2}}$$

once convergence is established. Here for $s\in\mathbb{C}$ and $u\in\mathbb{R}\backslash\{0\}$

$$\mathcal{W}_s(u):=|u|^{-\frac{1}{4}}\,W_{\frac{1}{4}\operatorname{sgn}(u),s-\frac{1}{2}}\left(|u|\right).$$

Proving convergence is quite involved and requires modifying a classical argument of Hooley [26]. To use his method, we rewrite the Kloosterman sums as Salie sums, which are sums over quadratic congruences. To be more precise, one can show that

$$\frac{A_{2k}\left(n-\frac{k\left(1+(-1)^k\right)}{4}\right)}{k}=\begin{cases}\frac{\rho_1(n;k)}{\sqrt{24k}} & \text{if } k \text{ is odd},\\ \frac{i\rho_1(n;k)}{\sqrt{24k}}-\frac{2ie\left(\frac{n}{2k}\right)\rho_2(n;k)}{\sqrt{24k}} & \text{if } k \text{ is even},\end{cases}$$

where

$$\rho_1(n;k):=\sum_{\substack{x \pmod{48k}\\ x^2\equiv 1-24n \pmod{48k}}}\chi_{12}(x)e\left(\frac{x}{24k}\right),$$
$$\rho_2(n;k):=\sum_{\substack{x \pmod{12k}\\ x^2\equiv 1-24n \pmod{12k}}}\chi_{12}(x)i^{\frac{x^2-1}{12k}}e\left(\frac{x}{24k}\right).$$

Using the fact that

$$I_{\frac{1}{2}}\left(\frac{\pi\sqrt{24n-1}}{12k}\right) \sim J_{\frac{1}{2}}\left(\frac{\pi\sqrt{|24n-1|}}{12k}\right) \sim \frac{|24n-1|^{\frac{1}{4}}}{\sqrt{6k}} \qquad (k \to \infty)$$

it is not hard to complete the proof once we establish the following estimates ($(j,\ell) \in \{(0,1),(0,2),(1,1)\}$):

$$\left|\sum_{\substack{k\geq 1\\ k\equiv j \pmod 2}} \frac{\rho_\ell(n;k)}{k}\right| \ll |24n-1|^{\frac{1}{2}}.$$

The proof of this estimate follows similarly as in Hooley. □

As the Poincaré series was constructed such that $\beta(n)$ agrees with the formula conjectured, we are left to show that

$$\alpha(n) = \beta(n).$$

For this we prove that the associated Maass forms coincide. The argument is somewhat lenghty and has been later simplified [13]. □

5 Mixed mock modular forms

We next turn to the Fourier coefficients of *mixed mock modular forms*, which are linear combinations of mock modular forms multiplied by modular forms. We note that there is no theory of such functions, the reason being that the space of harmonic Maass forms is not closed under multiplication (the eigenfunction property is violated). In particular, there are no corresponding Poincaré series known, and those were key ingredients for proving an exact formula for the Fourier coefficients of $f(q)$. To overcome these problems, we [10] developed an amplied version of the Hardy-Ramanujan Circle Method. Before describing the main modifications necessary, we want to give some examples.

The first example comes from combinatorics.

Definition. A *partition without sequences* [2, 32] is a partition without two consecutive numbers. We denote by $s(n)$ the number of partitions of n without sequences.

For example the partitions without sequences of 5 are given by

$$5, \quad 4+1, \quad 3+1+1, \quad 1+1+1+1+1.$$

Thus $s(5) = 4$.

We have the generating function [2]

$$G(q) := \sum_{n\geq 0} s(n)q^n = \frac{\left(-q^3;q^3\right)_\infty}{\left(q^2;q^2\right)_\infty} \cdot \chi(q),$$

where the product in front is a quotient of η-functions (and thus modular) and

$$\chi(q) := \sum_{n\geq 0} \frac{(-q;q)_n}{\left(-q^3;q^3\right)_n} q^{n^2}$$

is one of Ramanujan's third-order mock theta functions.

A further example comes from the generating function for Euler numbers of certain moduli spaces [11]. The functions that are of relevance here are ($j \in \{0,1\}$)

$$f_j(\tau) := \frac{1}{\eta^6(\tau)} \sum_{n\geq 0} H(4n+3j)q^{n+\frac{3j}{4}} =: \sum_{n\geq 0} \alpha_j(n) q^{n-\frac{j+1}{4}}. \tag{1.16}$$

The modularity of η and (1.13) yields that we have a mixed mock modular form.

A third example comes from Lie superalgebras. Recently, Kac and Wakimoto [28] found a character formula for certain $s\ell(m|1)^\wedge$-modules. In terms of q-series this can be written in the following way ($s \in \mathbb{Z}$):

$$2q^{-\frac{s}{2}} \cdot \frac{\left(q^2;q^2\right)_\infty^2}{(q;q)_\infty^{m+2}} \sum_{k=(k_1,k_2,\ldots,k_{m-1})\in\mathbb{Z}^{m-1}} \frac{q^{\frac{1}{2}\sum_{i=1}^{m-1} k_i(k_i+1)}}{1+q^{\sum_{i=1}^{m-1} k_i - s}}.$$

From this representation it is not clear at all that one has mixed mock modular forms, but the author proved this in joint work with Ken Ono [14].

One can also consider specialized character formulas for irreducible highest weight $s\ell(m|n)^\wedge$ modules. Here an even more complicated class of functions plays a role [8], the so-called *almost harmonic Maass forms*. Loosely speaking, these are sums of harmonic Maass forms under iterates of the raising operator $R_k := 2i\frac{\partial}{\partial\tau} + \frac{k}{y}$ (thus themselves non-harmonic weak Maass forms) multiplied by *almost holomorphic modular forms* [29], which are functions that transform like modular forms and are polynomials in $\frac{1}{y}$ with (weakly) holomorphic coefficients. We call the associated holomorphic parts *almost mock modular forms*.

There are many more interesting mixed mock modular forms, for example coming from certain generating function arising in the theory of black holes [18] or Joyce invariants.

As a special case, we treat the coefficients $\alpha_j(n)$ of f_j defined in (1.16). For this, we require some more notation. We let for $k \in \mathbb{N}$, $g \in \mathbb{Z}$, and $u \in \mathbb{R}$

$$f_{k,g}(u) := \begin{cases} \dfrac{\pi^2}{\sinh^2\left(\frac{\pi u}{k} - \frac{\pi i g}{2k}\right)} & \text{if } g \not\equiv 0 \pmod{2k}, \\ \dfrac{\pi^2}{\sinh^2\left(\frac{\pi u}{k}\right)} - \dfrac{k^2}{u^2} & \text{if } g \equiv 0 \pmod{2k}. \end{cases}$$

Furthermore, we define the Kloosterman sums

$$K_{j,\ell}(n,m;k) := \sum_{\substack{0\le h<k\\(h,k)=1}} \psi_{j\ell}(h,h',k)e^{-\frac{2\pi i}{k}\left(hn+\frac{h'm}{4}\right)},$$

where h' is defined by $hh' \equiv -1 \pmod{k}$ and where the $\psi_{j\ell}$ are certain explicit multipliers. Finally, we let

$$\mathcal{I}_{k,g}(n) := \int_{-1}^{1} f_{k,g}\left(\frac{u}{2}\right) I_{\frac{7}{2}}\left(\frac{\pi}{k}\sqrt{(4n-(j+1))\left(1-u^2\right)}\right)\left(1-u^2\right)^{\frac{7}{4}}du.$$

Theorem 5.1. *For $n \geq 1$, the Fourier coefficients $\alpha_j(n)$ of f_j are given by the following exact formula:*

$$\begin{aligned}\alpha_j(n) = &-\frac{\pi}{6}(4n-(j+1))^{-\frac{5}{4}}\sum_{k\ge1}\frac{K_{j,0}(n,0;k)}{k}I_{\frac{5}{2}}\left(\frac{\pi}{k}\sqrt{4n-(j+1)}\right)\\ &+\frac{1}{\sqrt{2}}(4n-(j+1))^{-\frac{3}{2}}\sum_{k\ge1}\frac{K_{j,0}(n,0;k)}{\sqrt{k}}I_3\left(\frac{\pi}{k}\sqrt{4n-(j+1)}\right)\\ &-\frac{1}{8\pi}(4n-(j+1))^{-\frac{7}{4}}\sum_{k\ge1}\sum_{\substack{\ell\in\{0,1\}\\-k<g\le k\\g\equiv\ell\ (\mathrm{mod}\ 2)}}\frac{K_{j,\ell}(n,g^2;k)}{k^2}\mathcal{I}_{k,g}(n).\end{aligned}$$

The above formula is useful as it allows one to determine asymptotic expansions. To be more precise, one can show (see [10, 11]) the following.

Corollary 5.2. *The leading asymptotic terms of $\alpha_j(n)$ for $n \to \infty$ are*

$$\alpha_j(n) = \left(\frac{1}{96}n^{-\frac{3}{2}} - \frac{1}{32\pi}n^{-\frac{7}{4}} + O\left(n^{-2}\right)\right)e^{2\pi\sqrt{n}}.$$

We note that in contrast to mock modular forms, the shadows of the mixed mock modular forms do contribute to the leading asymptotic terms. One could determine further polynomial lower order main terms.

Proof of Theorem 5.1. We only give a sketch of a proof. Details can be found in [11]. From (1.13) and the transformation law of the theta function it follows that

$$f_j\left(\frac{1}{k}(h+iz)\right) = z^{\frac{3}{2}}\sum_{\ell\in\{0,1\}}\psi_{j\ell}(h,h',k)e^{\frac{\pi i h'}{2k}-\frac{\pi i(j+1)h}{2k}}$$
$$\left(f_\ell\left(\frac{1}{k}\left(h'+\frac{i}{z}\right)\right)+\frac{1}{4\sqrt{2}\pi}\eta^{-6}\left(\frac{1}{k}\left(h'+\frac{i}{z}\right)\right)\mathcal{I}_\ell\left(\frac{1}{kz}\right)\right),$$

where

$$\mathcal{I}_j(x) := \int_0^\infty \frac{\Theta_j\left(iw - \frac{h'}{k}\right)}{(w+x)^{\frac{3}{2}}} dw.$$

Using partial fraction decomposition, we obtain

$$\mathcal{I}_j(x) = \sum_{\substack{g \pmod{2k} \\ g \equiv j \pmod{2}}} e\left(-\frac{g^2h'}{4k}\right)\left(\frac{2\delta_{0,g}}{\sqrt{x}} - \frac{1}{\sqrt{2}\pi k^2 x}\int_{-\infty}^{\infty} e^{-2\pi x u^2} f_{k,g}(u)du\right),$$

where $\delta_{0,g} = 0$ unless $g \equiv 0 \pmod{2k}$ in which case it equals 1.

Now the main difference compared to the use of the classical Circle Method is the contribution of integrals of the form

$$\mathcal{I}_{k,g,b}(z) := e^{\frac{2\pi b}{kz}} z^{\frac{5}{2}} \int_{-\infty}^{\infty} e^{-\frac{2\pi u^2}{kz}} f_{k,g}(u)du.$$

Similarly to the case of Fourier expansions, we are now interested in their “principal parts”. To be more precise, we let for $b > 0$ and $g \in \mathbb{Z}$

$$J_{k,g,b}(z) := e^{\frac{2\pi b}{kz}} z^{\frac{5}{2}} \int_{-\sqrt{b}}^{\sqrt{b}} e^{-\frac{2\pi u^2}{kz}} f_{k,g}(u)du.$$

One may show the following asymptotic bounds for $k < g \leq k$:

Lemma 5.3.

(i) *If $b \leq 0$, then*

$$\left|\mathcal{I}_{k,g,b}(z)\right| \ll |z|^{\frac{5}{2}} \begin{cases} \frac{k^2}{g^2} & \text{if } g \neq 0, \\ 1 & \text{if } g = 0. \end{cases}$$

(ii) *If $b > 0$, then*

$$\mathcal{I}_{k,g,b}(z) = J_{k,g,b}(z) + \mathcal{E}_{k,g,b}(z)$$

with $\mathcal{E}_{k,g,b}$ satisfying the bound in (i).

With this data one may then use the classical Circle Method. □

Acknowledgements

The author thanks Ben Kane and René Olivetto for useful comments on an earlier version.

Bibliography

[1] G. Andrews, On the theorems of Watson and Dragonette for Ramanujan's mock theta functions. *Am. J. Math.* 88 (1966), 454–490.

[2] G. Andrews, Partitions with short sequences and mock theta functions. *Proc. Natl. Acad. Sci. USA* 102 (2005), 4666–4671.

[3] G. Andrews, Partitions: At the interface of q-series and modular forms. *Ramanujan J.* 7 (2003), 385–400.

[4] G. Andrews, *The theory of partitions*. The Encyclopedia of Mathematics and its Applications series, Cambridge University Press (1998).

[5] G. Andrews and F. Garvan, Dyson's crank of a partition. *Bull. Am. Math. Soc.* 18 (1988), 167–171.

[6] A. Atkin and H. Swinnerton-Dyer, Some properties of partitions. *Proc. Lond. Math. Soc.* 4 (1954), 84–106.

[7] R. Borcherds, Automorphic forms with singularities on Grassmannians. *Invent. Math.* 132 (1998), 491–562.

[8] K. Bringmann and A. Folsom, Almost harmonic Maass forms and Kac–Wakimoto characters. *J. Reine Angew. Math.* (2014), 179–202.

[9] K. Bringmann and J. Lovejoy, Overpartitions and class numbers of binary quadratic forms. *Proc. Natl. Acad. Sci. USA* 106 (2009), 5513–5516.

[10] K. Bringmann and K. Mahlburg, An extension of the Hardy-Ramanujan Circle Method and applications to partitions without sequences. *Am. J. Math.* 133 (2011), 1151–1178.

[11] K. Bringmann and J. Manschot, From sheaves on $\mathbb{P}^2$ to generalisations of the Rademacher expansion. *Am. J. Math.* 135 (2013), 1039–1065.

[12] K. Bringmann and K. Ono, The $f(q)$ mock theta function conjecture and partition ranks. *Invent. Math.* 165 (2006), 243–266.

[13] K. Bringmann and K. Ono, Coefficients of harmonic weak Maass forms. *Proc. Conf. Part. Q-ser. Mod. Forms* (Univ. Florida), 2008.

[14] K. Bringmann and K. Ono, Some characters of Kac and Wakimoto and nonholomorphic modular functions. *Math. Ann.* 345 (2009), 547–558.

[15] J. Bruinier and J. Funke, On two geometric theta lifts. *Duke Math. J.* 125 (2004), 45–90.

[16] J. Bruinier and K. Ono, *Algebraic formulas for the coefficients of half-integral weight harmonic weak Maass forms*. *Advances in Mathematics* 246 (2013), 198–219.

[17] J. Bruinier, K. Ono and R. Rhoades, Differential operators and harmonic weak Maass forms. *Math. Ann.* 342 (2008), 673–693.

[18] A. Dabholkar, S. Murthy and D. Zagier, *Quantum black holes, wall crossing, and mock modular forms*. Preprint.

[19] L. Dragonette, Some asymptotic formulae for the mock theta series of Ramanujan. *Trans. Am. Math. Soc.* 72 (1952), 474–500.

[20] F. Dyson, Some guesses in the theory of partitions. *Eureka* (Cambridge) 8 (1944), 10–15.

[21] S. Garthwaite, Vector-valued Maass Poincaré series. *Proc. Am. Math. Soc.* 136 (2008), 427–436.

[22] G. Hardy and S. Ramanujan, Asymptotic formulae in combinatory analysis. *Proc. Lond. Math. Soc.*, 2, 17 (1918), 75–115.

[23] I. Gradshteyn and I. Ryzhik, *Tables of Integrals, Series, and Products*. Elsevier, Amsterdam 2007.

[24] D. Hejhal, *The Selberg trace formula for* $\mathrm{PSL}(2,\mathbb{R})$, part 2. Springer Lect. Notes in Math. 1001, Springer-Verlag, Berlin 1983.

[25] F. Hirzebruch and D. Zagier, Intersection numbers on curves on Hilbert modular surfaces and modular forms of Nebentypus. *Invent. Math.* 36 (1976), 57–113.

[26] C. Hooley, On the number of divisors of a quadratic polynomial. *Acta Math.* 110 (1963), 97–114.

[27] A. Ingham, A Tauberian theorem for partitions. *Ann. of Math.* 42 (1941), 1075–1090.

[28] V. Kac and M. Wakimoto, Integrable highest weight modules over affine superalgebras and Appell's function. *Comm. Math. Phys.* 215 (2011), 631–682.

[29] M. Kaneko and D. Zagier, A generalized Jacobi theta function and quasimodular forms. *The moduli space of curves, Progr. Math.* 129, Birkhäuser Boston, Massachusetts 1995, 165–172.

[30] M. Knopp, *Modular Functions in Analytic Number Theory*. Lectures in Advanced Mathematics, Markham Publishing Company, Chicago 1970.

[31] M. Köcher and A. Krieg, *Elliptische Funktionen und Modulformen*. Springer-Verlag, Berlin, 2007.

[32] P. MacMahon, *Combinatory Analysis*, Cambridge Univ. Press, Cambridge, U.K., Vol II, 49–58, 1915.

[33] M. Newman, Construction and application of a class of modular functions (II). *Proc. Lond. Math. Soc.* 9 (1959), 373–397.

[34] D. Niebur, A class of nonanalytic automorphic functions. *Nagoya Math. J.* 52 (1973), 133–145.

[35] D. Niebur, Construction of automorphic forms and integrals. *Trans. Am. Math. Soc.* 191 (1974), 373–385.

[36] H. Rademacher, On the expansion of the partition function in a series. *Ann. of Math.* 44 (1943), 416–422.

[37] S. Ramanujan, *The lost notebook and other unpublished papers*. Narosa, New Delhi 1988.

[38] G. Watson, The final problem: An account of the mock theta functions. *J. Lond. Math. Soc.* 11 (1936), 55–88.

[39] S. Zwegers, Mock ϑ-functions and real analytic modular forms. In *q-Ser. Appl. Comb. Number Theory Phys.* (ed. B. Berndt and K. Ono), Contemp. Math. 291, AMS (2001), 269–277.

Chapter 2

Expansions of algebraic numbers

Yann Bugeaud

Contents

1 Representation of real numbers

The most classical ways to represent real numbers are by means of their continued fraction expansion or their expansion in some integer base, in particular in base two or ten. In this text, we consider only these expansions and deliberately ignore β-expansions, Lüroth expansions, Q-Cantor series, etc., as well as the many variations of the continued fraction algorithm.

The first example of a transcendental number (recall that a real number is *algebraic* if it is a root of a nonzero polynomial with integer coefficients and it is *transcendental* otherwise) was given by Liouville [51, 52] in 1844. He showed that if the sequence of partial quotients of an irrational real number grows sufficiently rapidly, then this number is transcendental. He mentioned only at the very end of his note the now classical example of the series (keeping his notation)

$$\frac{1}{a} + \frac{1}{a^{1\cdot 2}} + \frac{1}{a^{1\cdot 2\cdot 3}} + \cdots + \frac{1}{a^{1\cdot 2\cdot 3\cdots m}} + \cdots,$$

where $a \geq 2$ is an integer.

Let b denote an integer at least equal to 2. Any real number ξ has a unique b-ary expansion, that is, it can be uniquely written as

$$\xi = \lfloor \xi \rfloor + \sum_{\ell \geq 1} \frac{a_\ell}{b^\ell} = \lfloor \xi \rfloor + 0 \cdot a_1 a_2 \ldots, \tag{2.1}$$

where $\lfloor \cdot \rfloor$ denotes the integer part function, the *digits* $a_1, a_2, \ldots$ are integers from the set $\{0, 1, \ldots, b-1\}$ and a_ℓ differs from $b-1$ for infinitely many indices ℓ. This notation will be kept throughout this text.

In a seminal paper published in 1909, Émile Borel [21] introduced the notion of *normal number*.

Definition 1.1. Let $b \geq 2$ be an integer. Let ξ be a real number whose b-ary expansion is given by (2.1). We say that ξ is normal to base b if, for every $k \geq 1$, every finite block of k digits in $\{0, 1, \ldots, b-1\}$ occurs with the same frequency $1/b^k$, that is, if for every $k \geq 1$ and every $d_1, \ldots, d_k \in \{0, 1, \ldots, b-1\}$ we have

$$\lim_{N \to +\infty} \frac{\#\{\ell : 0 \leq \ell < N, a_{\ell+1} = d_1, \ldots, a_{\ell+k} = d_k\}}{N} = \frac{1}{b^k}.$$

The above definition differs from that given by Borel, but is equivalent to it; see Chapter 4 of [27] for a proof and further equivalent definitions.

We reproduce the fundamental theorem proved by Borel in [21]. Throughout this text, ‘almost all’ always refers to the Lebesgue measure, unless otherwise specified.

Theorem 1.2. *Almost all real numbers are normal to every integer base $b \geq 2$.*

Despite the fact that normality is a property shared by almost all numbers, we do not know a single explicit example of a number normal to every integer base, let alone of a number normal to base 2 and to base 3. However, Martin [54] gave in 2001 a nice and simple explicit construction of a real number normal to no integer base.

The first explicit example of a real number normal to a given base was given by Champernowne [34] in 1933.

Theorem 1.3. *The real number*

$$0 \cdot 12345678910111213\ldots, \tag{2.2}$$

whose sequence of decimals is the increasing sequence of all positive integers, is normal to base ten.

Further examples also obtained by concatenation of sequences of integers have been given subsequently in [35, 37]. In particular, the real number

$$0 \cdot 235711131719232931\ldots, \tag{2.3}$$

whose sequence of decimals is the increasing sequence of all prime numbers, is normal to base ten. This is due to the fact that the sequence of prime numbers does not increase too rapidly. However, we still do not know whether the real numbers (2.2) and (2.3) are normal to base two.

Constructions of a completely different type were found by Stoneham [66] and Korobov [49]; see Bailey and Crandall [18] for a more general statement which includes the next theorem.

Theorem 1.4. *Let b and c be coprime integers, both at least equal to* 2*. Let $d \geq 2$ be an integer. Then, the real numbers*

$$\sum_{j\geq 1} \frac{1}{c^j b^{c^j}} \quad \textit{and} \quad \sum_{j\geq 1} \frac{1}{c^{d^j} b^{c^{d^j}}}$$

are normal to base b.

Regarding continued fraction expansions, we can as well define a notion of *normal continued fraction expansion* using the Gauss measure (see Section 5) and prove that the continued fraction expansion

$$\alpha = \lfloor\alpha\rfloor + [0; a_1, a_2, \ldots] = \lfloor\alpha\rfloor + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots}}}$$

of almost every real number α is a normal continued fraction expansion. Here, the positive integers $a_1, a_2, \ldots$ are called the *partial quotients* of α (throughout this text, a_ℓ denotes either a b-ary digit or a partial quotient, but this should be clear from the context; furthermore, we write ξ for a real number when we consider its b-ary expansion and α when we study its continued fraction expansion). In 1981, Adler, Keane, and Smorodinsky [13] have constructed a normal

continued fraction in a similar way as Champernowne did for a normal number.

Theorem 1.5. *Let $1/2, 1/3, 2/3, 1/4, 2/4, 3/4, \ldots$ be the infinite sequence obtained in writing the rational numbers in $(0, 1)$ with denominator* 2*, then with denominator* 3*, denominator* 4*, etc., ordered with numerators increasing. Let $x_1x_2x_3\ldots$ be the sequence of positive integers constructed by concatenating the partial quotients* (*we choose the continued fraction expansion which does not end with the digit* 1) *of this sequence of rational numbers. Then, the real number*

$$[0; x_1, x_2, \ldots] = [0; 2, 3, 1, 2, 4, 2, 1, 3, 5, \ldots]$$

has a normal continued fraction expansion.

All this shows that the b-ary expansion and the continued fraction expansion of a real number taken at random are well understood. But what can be said for a specific number, like $\sqrt[3]{2}, \log 2, \pi$, etc.?

Actually, not much! We focus our attention on algebraic numbers. Clearly, a real number is rational if, and only if, its b-ary expansion is ultimately periodic. Analogously, a real number is quadratic if, and only if, its continued fraction expansion is ultimately periodic; see Section 5. The purpose of the present text is to gather what is known on the b-ary expansion of an irrational algebraic number and on the continued fraction expansion of an algebraic number of degree at least three. It is generally believed that all these expansions are normal, and some numerical computation tend to support this guess, but we are very, very far from proving such a strong assertion. We still do not know whether there is an integer $b \geq 3$ such that at least three different digits occur infinitely often in the b-ary expansion of $\sqrt{2}$. And whether there exist algebraic numbers of degree at least three whose sequence of partial quotients is bounded. Actually, it is widely believed that algebraic numbers should share most of the properties of almost all real numbers. This is indeed the case from the point of view of rational approximation, since Roth's theorem (see Section 6) asserts that algebraic irrational numbers do behave like almost all numbers, in the sense that they cannot be approximated by rational numbers at an order greater than 2.

The present text is organized as follows. Section 2 contains basic results from combinatorics on words. The main results on the complexity of algebraic numbers are stated in Section 3. They are proved by combining combinatorial transcendence criteria given in Section 4 and established in Sections 9 and 10 with a combinatorial lemma proved in Section 8. In Sections 5 and 6 we present various auxiliary results from the theory of continued fractions and from Diophantine approximation, respectively. Section 7 is devoted to a sketch of the proof of Theorem 4.1 and to a short historical discussion. We present in Section 11 another combinatorial transcendence criterion for continued fraction expansions, along with its proof. Section 12 briefly surveys some refined results which complement Theorem 3.1. Finally, in Section 13, we discuss other points of view for measuring the complexity of the b-ary expansion of a number.

A proof of Theorem 4.1 can already be found in the surveys [20, 8] and in the monograph [27]. Here, we provide two different proofs. Historical remarks and discussion on the various results which have ultimately led to Theorem 4.2 are given in [30].

2 Combinatorics on words and complexity

In the sequel, we often identify a real number with the infinite sequence of its b-ary digits or of its partial quotients. It appears to be convenient to use the point of view from combinatorics on words. Throughout, we denote by $\mathcal{A}$ a finite or infinite set. A finite word over the alphabet $\mathcal{A}$ is either the empty word, or a finite string (or block) of elements from $\mathcal{A}$. An infinite word over $\mathcal{A}$ is an infinite sequence of elements from $\mathcal{A}$.

For an infinite word $\mathbf{w} = w_1 w_2 \dots$ over the alphabet $\mathcal{A}$ and for any positive integer n, we let

$$p(n, \mathbf{w}, \mathcal{A}) := \#\{w_{j+1} \dots w_{j+n} : j \geq 0\}$$

denote the number of distinct strings (or blocks) of length n occurring in $\mathbf{w}$. Obviously, putting $\#\mathcal{A} = +\infty$ if $\mathcal{A}$ is infinite, we have

$$1 \leq p(n, \mathbf{w}, \mathcal{A}) \leq (\#\mathcal{A})^n,$$

and both inequalities are sharp. Furthermore, the function $n \mapsto p(n, \mathbf{w}, \mathcal{A})$ is non-decreasing.

Definition 2.1. An infinite word $\mathbf{w} = w_1 w_2 \dots$ is ultimately periodic if there exist positive integers n_0 and T such that

$$w_{n+T} = w_n, \quad \text{for every } n \geq n_0.$$

The word $w_{n_0} w_{n_0+1} \dots w_{n_0+T-1}$ is a period of $\mathbf{w}$. If n_0 can be chosen equal to 1, then $\mathbf{w}$ is (purely) periodic, otherwise, $w_1 \dots w_{n_0-1}$ is a preperiod of $\mathbf{w}$.

We establish a seminal result from Morse and Hedlund [55, 56].

Theorem 2.2. *Let* $\mathbf{w}$ *be an infinite word over a finite or infinite alphabet* $\mathcal{A}$. *If* $\mathbf{w}$ *is ultimately periodic, then there exists a positive constant* C *such that* $p(n, \mathbf{w}, \mathcal{A}) \leq C$ *for every positive integer* n. *Otherwise, we have*

$$p(n+1, \mathbf{w}, \mathcal{A}) \geq p(n, \mathbf{w}, \mathcal{A}) + 1 \quad \textit{for every } n \geq 1,$$

thus,

$$p(n, \mathbf{w}, \mathcal{A}) \geq n + 1 \quad \textit{for every } n \geq 1.$$

Proof. Throughout the proof, we write $p(\cdot, \mathbf{w})$ instead of $p(\cdot, \mathbf{w}, \mathcal{A})$.

Let $\mathbf{w}$ be an ultimately periodic infinite word, and assume that it has a preperiod of length r and a period of length s. Fix $h = 1, \dots, s$ and let n be a positive integer. For every $j \geq 1$, the block of length n starting at w_{r+js+h} is the same as the one starting at w_{r+h}. Consequently, there cannot be more than $r + s$ distinct blocks of length n, thus, $p(n, \mathbf{w}) \leq r + s$.

Write $\mathbf{w} = w_1 w_2 \dots$ and assume that there is a positive integer n_0 such that $p(n_0, \mathbf{w}) = p(n_0 + 1, \mathbf{w})$. This means that every block of length n_0 extends uniquely to a block of length $n_0 + 1$. It implies that $p(n_0, \mathbf{w}) = p(n_0 + j, \mathbf{w})$ holds for every positive integer j. By the definition of $p(n_0, \mathbf{w})$, two among the words $w_j \dots w_{n_0+j-1}$, $j = 1, \dots, p(n_0, \mathbf{w}) + 1$, are the same. Consequently, there are integers k and ℓ with $0 \leq k < \ell \leq p(n_0, \mathbf{w})$ and $w_{k+m} = w_{\ell+m}$ for $m = 1, \dots, n_0$. Since every block of length n_0 extends uniquely to a block of length $n_0 + 1$, this gives $w_{k+m} = w_{\ell+m}$ for every positive integer m. This proves that the word $\mathbf{w}$ is ultimately periodic.

Consequently, if $\mathbf{w}$ is not ultimately periodic, then $p(n+1, \mathbf{w}) \geq p(n, \mathbf{w}) + 1$ holds for every positive integer n. Then, $p(1, \mathbf{w}) \geq 2$ and an immediate induction show that $p(n, \mathbf{w}) \geq n + 1$ for every n. The proof of the theorem is complete. □

We complement Theorem 2.2 by pointing out that there exist uncountably many infinite words $\mathbf{w}$ over $\mathcal{A} = \{0, 1\}$ such that

$$p(n, \mathbf{w}, \mathcal{A}) = n + 1, \quad \text{for } n \geq 1.$$

These words are called *Sturmian words*; see e.g. [16].

To prove that a real number is normal to some given integer base, or has a normal continued fraction expansion, is in most cases a far too difficult problem. So we are led to consider weaker questions on the sequence of digits (resp. partial quotients), including the following ones:

- Does every digit occur infinitely many times in the b-ary expansion of ξ?
- Are there many non-zero digits in the b-ary expansion of ξ?
- Is the sequence of partial quotients of α bounded from above?
- Does the sequence of partial quotients of α tend to infinity?

We may even consider weaker questions, namely, and this is the point of view we adopt until the very last sections, we wish to bound from below the number of different blocks in the infinite word composed of the digits of ξ (resp. partial quotients of α).

Let $b \geq 2$ be an integer. A natural way to measure the *complexity* of a real number ξ whose b-ary expansion is given by (2.1) is to count the number of distinct blocks of given length in the infinite word $\mathbf{a} = a_1 a_2 a_3 \ldots$ We set $p(n, \xi, b) = p(n, \mathbf{a}, b)$, with $\mathbf{a}$ as above. Clearly, we have

$$p(n, \xi, b) = \#\{a_{j+1} a_{j+2} \ldots a_{j+n} : j \geq 0\} = p(n, \mathbf{a}, \{0, 1, \ldots, b-1\})$$

and

$$1 \leq p(n, \xi, b) \leq b^n,$$

where both inequalities are sharp.

Since the b-ary expansion of a real number is ultimately periodic if, and only if, this number is rational, Theorem 2.2 can be restated as follows.

Theorem 2.3. *Let $b \geq 2$ be an integer. If the real number ξ is irrational, then*

$$p(n, \xi, b) \geq n + 1, \quad \textit{for } n \geq 1.$$

Otherwise, the sequence $(p(n, \xi, b))_{n \geq 1}$ is bounded.

Let α be an irrational real number and write

$$\alpha = \lfloor \alpha \rfloor + [0; a_1, a_2, \ldots].$$

Let **a** denote the infinite word $a_1a_2\dots$ over the alphabet $\mathbb{Z}_{\geq 1}$. A natural way to measure the intrinsic *complexity* of α is to count the number $p(n,\alpha) := p(n,\mathbf{a},\mathbb{Z}_{\geq 1})$ of distinct blocks of given length n in the word **a**.

Since the continued fraction expansion of a real number is ultimately periodic if, and only if, this number is quadratic (see Theorem 5.7), Theorem 2.2 can be restated as follows.

Theorem 2.4. *Let $b \geq 2$ be an integer. If the real number α is irrational and not quadratic, then*

$$p(n,\alpha) \geq n+1, \quad \textit{for } n \geq 1.$$

If the real number α is quadratic, then the sequence $(p(n,\alpha))_{n\geq 1}$ is bounded.

We show in the next section that Theorem 2.3 (resp. 2.4) can be improved when ξ (resp. α) is assumed to be algebraic.

3 Complexity of algebraic numbers

As already mentioned, we focus on the digital expansions and on the continued fraction expansion of algebraic numbers. Until the end of the 20th century, it was only known that the sequence of partial quotients of an algebraic number cannot grow too rapidly; see Section 12. Regarding b-ary expansions, Ferenczi and Mauduit [43] were the first to improve the (trivial) lower bound given by Theorem 2.3 for the complexity function of the b-ary expansion of an irrational

algebraic number θ. They showed in 1997 that $p(n,\theta,b)$ strictly exceeds $n+1$ for every sufficiently large integer n. Actually, as pointed out a few years later by Allouche [14], their approach combined with a combinatorial result of Cassaigne [32] yields a slightly stronger result, namely that

$$\lim_{n\to+\infty}\big(p(n,\theta,b)-n\big) = +\infty, \tag{2.4}$$

for any algebraic irrational number θ.

The estimate (2.4) follows from a good understanding of the combinatorial structure of Sturmian sequences combined with a combinatorial translation of Ridout's theorem 6.6. The transcendence criterion given in Theorem 4.1, established in [10, 3], yields an improvement of (2.4).

Theorem 3.1. *For any irrational algebraic number θ and any integer $b \geq 2$, we have*

$$\lim_{n\to+\infty}\frac{p(n,\theta,b)}{n} = +\infty. \tag{2.5}$$

Although (2.5) considerably strengthens (2.4), it is still very far from what is commonly expected, that is, from confirming that $p(n,\theta,b) = b^n$ holds for every positive n when θ is algebraic irrational.

Regarding continued fraction expansions, it was proved in [15] that

$$\lim_{n\to+\infty}\big(p(n,\theta)-n\big)=+\infty,$$

for any algebraic number θ of degree at least three. This is the continued fraction analogue of (2.4).

Using ideas from [1], the continued fraction analogue of Theorem 3.1 was established in [29].

Theorem 3.2. *For any algebraic number θ of degree at least three, we have*

$$\lim_{n\to+\infty}\frac{p(n,\theta)}{n}=+\infty. \tag{2.6}$$

The main purpose of the present text is to give complete (if one admits Theorem 6.7, whose proof is much too long and involved to be included here) proofs of Theorems 3.1 and 3.2. They are established by combining combinatorial transcendence criteria (Theorems 4.1 and 4.2) and a combinatorial lemma (Lemma 8.1). This is explained in details at the end of Section 8.

4 Combinatorial transcendence criteria

In this section, we state the combinatorial transcendence criteria which, combined with the combinatorial lemma from Section 8, yield Theorems 3.1 and 3.2.

Throughout, the length of a finite word W over the alphabet $\mathcal{A}$, that is, the number of letters composing W, is denoted by $|W|$.

Let $\mathbf{a}=(a_\ell)_{\ell\ge1}$ be a sequence of elements from $\mathcal{A}$. We say that $\mathbf{a}$ satisfies Condition (♠) if $\mathbf{a}$ is not

ultimately periodic and if there exist three sequences of finite words, $(U_n)_{n\ge1}$, $(V_n)_{n\ge1}$, and $(W_n)_{n\ge1}$, such that:

(i) For every $n\ge1$, the word $W_nU_nV_nU_n$ is a prefix of the word $\mathbf{a}$.

(ii) The sequence $(|V_n|/|U_n|)_{n\ge1}$ is bounded from above.

(iii) The sequence $(|W_n|/|U_n|)_{n\ge1}$ is bounded from above.

(iv) The sequence $(|U_n|)_{n\ge1}$ is increasing.

Theorem 4.1. *Let $b\ge2$ be an integer. Let $\mathbf{a}=(a_\ell)_{\ell\ge1}$ be a sequence of elements from $\{0,1,\dots,b-1\}$. If $\mathbf{a}$ satisfies Condition (♠), then the real number*

$$\xi:=\sum_{\ell=1}^{+\infty}\frac{a_\ell}{b^\ell}$$

is transcendental.

Theorem 4.2. *Let* $\mathbf{a} = (a_\ell)_{\ell \geq 1}$ *be a sequence of positive integers. Let* $(p_\ell/q_\ell)_{\ell \geq 1}$ *denote the sequence of convergents to the real number*

$$\alpha := [0; a_1, a_2, \ldots, a_\ell, \ldots].$$

Assume that the sequence $(q_\ell^{1/\ell})_{\ell \geq 1}$ *is bounded. If* **a** *satisfies Condition* (♠)*, then* α *is transcendental.*

The common tool for the proofs of Theorems 4.1 and 4.2 is a powerful theorem from Diophantine approximation theory, the Subspace Theorem; see Section 6.

Let us comment on Condition (♠) when, for simplicity, the alphabet $\mathcal{A}$ is finite and has $b \geq 2$ elements. Take an arbitrary infinite word $a_1 a_2 \ldots$ on $\{0, 1, \ldots, b-1\}$. Then, by the *Schubfachprinzip*, for every positive integer m, there exists (at least) one finite word U_m of length m having (at least) two (possibly overlapping) occurrences in the prefix $a_1 a_2 \ldots a_{b^m+m}$. If, for simplicity, we suppose that these two occurrences do not overlap, then there exist finite (or empty) words V_m, W_m, X_m such that

$$a_1 a_2 \ldots a_{b^m+m} = W_m U_m V_m U_m X_m.$$

This simple argument gives no additional information on the lengths of W_m and V_m, which a priori can be as large as $b^m - m$. In particular, they can be larger than some constant greater than 1 raised to the power the length of U_m.

We demand much more for a sequence **a** to satisfy Condition (♠), namely we impose that there exists an integer C such that, for infinitely many m, the lengths of V_m and W_m do not exceed C times the length of U_m. Such a condition occurs quite rarely.

We end this section with a few comments on de Bruijn words.

Definition 4.3. Let $b \geq 2$ and $n \geq 1$ be integers. A de Bruijn word of order n over an alphabet of cardinality b is a word of length $b^n + n - 1$ in which every block of length n occurs exactly once.

A recent result of Becher and Heiber [19] shows that we can extend de Bruijn words.

Theorem 4.4. *Every de Bruijn word of order* n *over an alphabet with at least three letters can be extended to a de Bruijn word of order* $n+1$*. Every de Bruijn word of order* n *over an alphabet with two letters can be extended to a de Bruijn word of order* $n+2$.

Theorem 4.4 shows that there exist infinite de Bruijn words obtained as the inductive limit of extended de Bruijn sequences of order n, for each n (when the alphabet has at least three letters; for each even n, otherwise). Let $b \geq 2$ be an integer. By construction, for every $m \geq 1$, the shortest prefix of an infinite de Bruijn word having two occurrences of a same word of length m has at least $b^m + m$ letters if $b \geq 3$ and at least $2^{m-1} + m - 1$ letters if $b = 2$.

5 Continued fractions

In this section, we briefly present classical results on continued fractions which will be used in the proofs of Theorems 4.2 and 11.1. We omit most of the proofs and refer the reader to a text of van der Poorten [58] and to the books of Bugeaud [22], Cassels [33], Hardy and Wright [44], Khintchine [47], Perron [57], and Schmidt [64], among many others.

Let $x_0, x_1, \dots$ be real numbers with $x_1, x_2, \dots$ positive. A *finite continued fraction* denotes any expression of the form

$$[x_0; x_1, x_2, \dots, x_n] = x_0 + \cfrac{1}{x_1 + \cfrac{1}{x_2 + \cfrac{1}{\dots + \cfrac{1}{x_n}}}}.$$

We call any expression of the above form or of the form

$$[x_0; x_1, x_2, \dots] = x_0 + \cfrac{1}{x_1 + \cfrac{1}{x_2 + \cfrac{1}{\dots}}} = \lim_{n \to +\infty} [x_0; x_1, x_2, \dots, x_n]$$

a *continued fraction*, provided that the limit exists.

Any rational number r has exactly two different continued fraction expansions. These are $[r]$ and $[r-1;1]$ if r is an integer and, otherwise, one of them reads $[a_0; a_1, \dots, a_{n-1}, a_n]$ with $a_n \ge 2$, and the other one is $[a_0; a_1, \dots, a_{n-1}, a_n - 1, 1]$. Any irrational number has a unique expansion in continued fraction.

Theorem 5.1. *Let $\alpha = [a_0; a_1, a_2, \dots]$ be an irrational number. For $\ell \ge 1$, set $p_\ell / q_\ell := [a_0; a_1, a_2, \dots, a_\ell]$. Let n be a positive integer. Putting*

$$p_{-1} = 1, \quad q_{-1} = 0, \quad p_0 = a_0, \quad \textit{and} \quad q_0 = 1,$$

we have

$$p_n = a_n p_{n-1} + p_{n-2}, \quad q_n = a_n q_{n-1} + q_{n-2}, \tag{2.7}$$

and

$$p_{n-1} q_n - p_n q_{n-1} = (-1)^n. \tag{2.8}$$

Furthermore, setting $\alpha_{n+1} = [a_{n+1}; a_{n+2}, a_{n+3}, \dots]$, we have

$$\alpha = [a_0; a_1, \dots, a_n, \alpha_{n+1}] = \frac{p_n \alpha_{n+1} + p_{n-1}}{q_n \alpha_{n+1} + q_{n-1}}, \tag{2.9}$$

thus

$$q_n \alpha - p_n = \frac{(-1)^n}{q_n \alpha_{n+1} + q_{n-1}},$$

and

$$\frac{1}{(a_{n+1}+2)q_n^2} < \frac{1}{q_n(q_n+q_{n+1})} < \left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_n q_{n+1}} < \frac{1}{a_{n+1}q_n^2} \le \frac{1}{q_n^2}. \quad (2.10)$$

It follows from (2.9) that any real number whose first partial quotients are $a_0, a_1, \ldots, a_n$ belongs to the interval with endpoints $(p_n + p_{n-1})/(q_n + q_{n-1})$ and p_n/q_n. Consequently, we get from (2.8) an upper bound for the distance between two real numbers having the same first partial quotients.

Corollary 5.2. *Let $\alpha = [a_0; a_1, a_2, \ldots]$ be an irrational number. For $\ell \ge 0$, let q_ℓ be the denominator of the rational number $[a_0; a_1, a_2, \ldots, a_\ell]$. Let n be a positive integer and β be a real number such that the first partial quotients of β are $a_0, a_1, \ldots, a_n$. Then,*

$$|\alpha - \beta| \le \frac{1}{q_n(q_n + q_{n-1})} < \frac{1}{q_n^2}.$$

Under the assumption of Theorem 5.1, the rational number p_ℓ/q_ℓ is called the *ℓ-th convergent to α*. It follows from (2.7) that the sequence of denominators of convergents grows at least exponentially fast.

Theorem 5.3. *Let $\alpha = [a_0; a_1, a_2, \ldots]$ be an irrational number. For $\ell \ge 0$, let q_ℓ be the denominator of the rational number $[a_0; a_1, a_2, \ldots, a_\ell]$. For any positive integers ℓ, h, we have*

$$q_{\ell+h} \ge q_\ell (\sqrt{2})^{h-1}$$

and

$$q_\ell \le (1 + \max\{a_1, \ldots, a_\ell\})^\ell.$$

Proof. The first assertion follows via induction on h, since $q_{n+2} \ge q_{n+1} + q_n \ge 2q_n$ for every $n \ge 0$. The second assertion is an immediate consequence of (2.7). □

The next result is sometimes called the *mirror formula*.

Theorem 5.4. *Let $n \ge 2$ be an integer and $a_1, \ldots, a_n$ be positive integers. For $\ell = 1, \ldots, n$, set $p_\ell/q_\ell = [0; a_1, \ldots, a_\ell]$. Then,*

$$\frac{q_{n-1}}{q_n} = [0; a_n, a_{n-1}, \ldots, a_1].$$

Proof. We get from (2.7) that

$$\frac{q_n}{q_{n-1}} = a_n + \frac{q_{n-2}}{q_{n-1}},$$

for $n \ge 1$. The theorem then follows by induction. □

An alternative proof of Theorem 5.4 goes as follows. Observe that, if $a_0 = 0$ and $n \ge 1$, then, by (2.7),

$$M_n := \begin{pmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_2 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix}.$$

Taking the transpose, we immediately get that

$$\begin{aligned} {}^tM_n &= {}^t\left(\begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_2 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix}\right) \\ &= {}^t\begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix} {}^t\begin{pmatrix} 0 & 1 \\ 1 & a_{n-1} \end{pmatrix} \cdots {}^t\begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_{n-1} \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} = \begin{pmatrix} p_{n-1} & q_{n-1} \\ p_n & q_n \end{pmatrix}, \end{aligned}$$

which gives Theorem 5.4.

Theorem 5.4 is a particular case of a more general result, which we state below after introducing the notion of *continuant*.

Definition 5.5. Let $m \ge 1$ and $a_1, \ldots, a_m$ be positive integers. The denominator of the rational number $[0; a_1, \ldots, a_m]$ is called the continuant of $a_1, \ldots, a_m$ and is usually denoted by $K_m(a_1, \ldots, a_m)$.

Theorem 5.6. *For any positive integers $a_1, \ldots, a_m$ and any integer k with $1 \le k \le m-1$, we have*

$$K_m(a_1, \ldots, a_m) = K_m(a_m, \ldots, a_1), \tag{2.11}$$

and

$$\begin{aligned} K_k(a_1, \ldots, a_k) \cdot K_{m-k}(a_{k+1}, \ldots, a_m) & \\ \le K_m(a_1, \ldots, a_m) & \\ \le 2\, K_k(a_1, \ldots, a_k) \cdot K_{m-k}(a_{k+1}, \ldots, a_m). & \end{aligned} \tag{2.12}$$

Proof. The first statement is an immediate consequence of Theorem 5.4. Combining

$$K_m(a_1, \ldots, a_m) = a_m K_{m-1}(a_1, \ldots, a_{m-1}) + K_{m-2}(a_1, \ldots, a_{m-2})$$

with (2.11), we get

$$K_m(a_1, \ldots, a_m) = a_1 K_{m-1}(a_2, \ldots, a_m) + K_{m-2}(a_3, \ldots, a_m),$$

which implies (2.12) for $k = 1$. Let $k \in \{1, 2, \ldots, m-2\}$ be such that

$$\begin{aligned} K_m &:= K_m(a_1, \ldots, a_m) \\ &= K_k(a_1, \ldots, a_k) \cdot K_{m-k}(a_{k+1}, \ldots, a_m) \\ &\quad + K_{k-1}(a_1, \ldots, a_{k-1}) \cdot K_{m-k-1}(a_{k+2}, \ldots, a_m), \end{aligned} \tag{2.13}$$

where we have set $K_0 = 1$. We then have

$$\begin{aligned} K_m &= K_k(a_1,\dots,a_k)\cdot\big(a_{k+1}K_{m-k-1}(a_{k+2},\dots,a_m) + K_{m-k-2}(a_{k+3},\dots,a_m)\big)\\ &\quad + K_{k-1}(a_1,\dots,a_{k-1})\cdot K_{m-k-1}(a_{k+2},\dots,a_m)\\ &= \big(a_{k+1}K_k(a_1,\dots,a_k) + K_{k-1}(a_1,\dots,a_{k-1})\big)\cdot K_{m-k-1}(a_{k+2},\dots,a_m)\\ &\quad + K_k(a_1,\dots,a_k)\cdot K_{m-k-2}(a_{k+3},\dots,a_m), \end{aligned}$$

giving (2.13) for the index $k+1$. This shows that (2.13) and, a fortiori, (2.12) hold for $k = 1,\dots,m-1$. □

The ‘only if’ part of the next theorem is due to Euler [39], and the ‘if’ part was established by Lagrange [50] in 1770.

Theorem 5.7. *The real irrational number $\alpha = [a_0; a_1, a_2, \dots]$ has a periodic continued fraction expansion (that is, there exist integers $r \ge 0$ and $s \ge 1$ such that $a_{n+s} = a_n$ for all integers $n \ge r+1$) if, and only if, α is a quadratic irrationality.*

We display an elementary result on ultimately periodic continued fraction expansions.

Lemma 5.8. *Let θ be a quadratic real number with ultimately periodic continued fraction expansion*

$$\theta = [0; a_1,\dots,a_r,\overline{a_{r+1},\dots,a_{r+s}}],$$

and denote by $(p_\ell/q_\ell)_{\ell\ge1}$ the sequence of its convergents. Then, θ is a root of the polynomial

$$\begin{aligned}(q_{r-1}q_{r+s} - q_rq_{r+s-1})X^2 - (q_{r-1}p_{r+s} - q_rp_{r+s-1} + p_{r-1}q_{r+s} - p_rq_{r+s-1})X\\ + (p_{r-1}p_{r+s} - p_rp_{r+s-1}).\end{aligned} \tag{2.14}$$

Proof. It follows from (2.9) that

$$\theta = [0; a_1,\dots,a_r,\theta_{r+1}] = \frac{p_r\theta' + p_{r-1}}{q_r\theta' + q_{r-1}} = \frac{p_{r+s}\theta' + p_{r+s-1}}{q_{r+s}\theta' + q_{r+s-1}},$$

where $\theta' = [a_{r+1}; \overline{a_{r+2},\dots,a_{r+s},a_{r+1}}]$. Consequently, we get

$$\theta' = \frac{p_{r-1} - q_{r-1}\theta}{q_r\theta - p_r} = \frac{p_{r+s-1} - q_{r+s-1}\theta}{q_{r+s}\theta - p_{r+s}},$$

from which we obtain

$$(p_{r-1} - q_{r-1}\theta)(q_{r+s}\theta - p_{r+s}) = (p_{r+s-1} - q_{r+s-1}\theta)(q_r\theta - p_r).$$

This shows that θ is a root of (2.14). □

We do not claim that the polynomial in (2.14) is the minimal polynomial of θ over the integers. This is indeed not always true, since its coefficients may have common prime factors.

The sequence of partial quotients of an irrational real number α in $(0, 1)$ can be obtained by iterations of the Gauss map $T_{\mathcal{G}}$ defined by $T_{\mathcal{G}}(0) = 0$ and $T_{\mathcal{G}}(x) = \{1/x\}$ for $x \in (0, 1)$. Namely, if $[0; a_1, a_2, \ldots]$ denotes the continued fraction expansion of α, then $T_{\mathcal{G}}^n(\alpha) = [0; a_{n+1}, a_{n+2}, \ldots]$ and $a_n = \lfloor 1/T_{\mathcal{G}}^{n-1}(\alpha) \rfloor$ for $n \geq 1$.

In the sequel, it is understood that α is a real number in $(0, 1)$, whose partial quotients $a_1(\alpha), a_2(\alpha), \ldots$ and convergents $p_1(\alpha)/q_1(\alpha), p_2(\alpha)/q_2(\alpha), \ldots$ are written $a_1, a_2, \ldots$ and $p_1/q_1, p_2/q_2, \ldots$, respectively, when there is no danger of confusion.

The map $T_{\mathcal{G}}$ possesses an invariant ergodic probability measure, namely the Gauss measure $\mu_{\mathcal{G}}$, which is absolutely continuous with respect to the Lebesgue measure, with density

$$\mu_{\mathcal{G}}(\mathrm{d}x) = \frac{\mathrm{d}x}{(1+x)\log 2}.$$

For every function f in $L^1(\mu_{\mathcal{G}})$ and almost every α in $(0, 1)$, we have (Theorem 3.5.1 in [36])

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T_{\mathcal{G}}^k \alpha) = \frac{1}{\log 2} \int_0^1 \frac{f(x)}{1+x} \, \mathrm{d}x. \tag{2.15}$$

For subsequent results in the metric theory of continued fractions, we refer the reader to [47] and to [36].

Definition 5.9. We say that $[0; a_1, a_2, \ldots]$ is a normal continued fraction, if, for every integer $k \geq 1$ and every positive integers $d_1, \ldots, d_k$, we have

$$\begin{aligned} &\lim_{N \to +\infty} \frac{\#\{j : 0 \leq j \leq N-k, a_{j+1} = d_1, \ldots, a_{j+k} = d_k\}}{N} \\ &\quad = \int_{r/s}^{r'/s'} \mu_{\mathcal{G}}(\mathrm{d}x) = \mu_{\mathcal{G}}(\Delta_{d_1,\ldots,d_k}), \end{aligned} \tag{2.16}$$

where r/s and r'/s' denote the rational numbers $[0; d_1, \ldots, d_{k-1}, d_k]$ and $[0; d_1, \ldots, d_{k-1}, d_k + 1]$, ordered so that $r/s < r'/s'$, and $\Delta_{d_1,\ldots,d_k} = [r/s, r'/s']$.

Let $d_1, \ldots, d_k$ be positive integers. It follows from Theorem 5.1 that the set $\Delta_{d_1,\ldots,d_k}$ of real numbers α in $(0, 1)$ whose first k partial quotients are $d_1, \ldots, d_k$ is an interval of length $1/(q_k(q_k + q_{k-1}))$. Applying (2.15) to the function $f = \mathbf{1}_{\Delta_{d_1,\ldots,d_k}}$, we get that for almost every $\alpha = [0; a_1, a_2, \ldots]$ in $(0, 1)$ the limit defined in (2.16) exists and is equal to $\mu_{\mathcal{G}}(\Delta_{d_1,\ldots,d_k})$. Thus, we have established the following statement.

Theorem 5.10. *Almost every α in $(0, 1)$ has a normal continued fraction expansion.*

The construction of [13], reproduced in [27], is flexible enough to produce many examples of real numbers with a normal continued fraction expansion.

6 Diophantine approximation

In this section, we survey classical results on approximation of real (algebraic) numbers by rational numbers.

We emphasize one of the consequences of Theorem 5.1.

Theorem 6.1. *For every real irrational number ξ, there exist infinitely many rational numbers p/q with $q \ge 1$ and*

$$\left| \xi - \frac{p}{q} \right| < \frac{1}{q^2}.$$

Theorem 6.1 is often, and wrongly, attributed to Dirichlet, who proved in 1842 a stronger result, namely that, under the assumption of Theorem 6.1 and for every integer $Q \ge 1$, there exist integers p, q with $1 \le q \le Q$ and $|\xi - p/q| < 1/(qQ)$. Theorem 6.1 was proved long before 1842.

An easy covering argument shows that, for almost all numbers, the exponent of q in Theorem 6.1 cannot be improved.

Theorem 6.2. *For every $\varepsilon > 0$ and almost all real numbers ξ, there exist only finitely many rational numbers p/q with $q \ge 1$ and*

$$\left| \xi - \frac{p}{q} \right| < \frac{1}{q^{2+\varepsilon}}. \tag{2.17}$$

Proof. Without loss of generality, we may assume that ξ is in $(0,1)$. If there are infinitely many rational numbers p/q with $q \ge 1$ satisfying (2.17), then ξ belongs to the limsup set

$$\bigcap_{Q \ge 1} \bigcup_{q \ge Q} \bigcup_{p=0}^{q} \left(\frac{p}{q} - \frac{1}{q^{2+\varepsilon}}, \frac{p}{q} + \frac{1}{q^{2+\varepsilon}} \right) \cap (0,1).$$

The Lebesgue measure of the latter set is, for every $Q \ge 1$, at most equal to

$$\sum_{q \ge Q} q \, \frac{2}{q^{2+\varepsilon}},$$

which is the tail of a convergent series and thus tends to 0 as Q tends to infinity. This proves the theorem. □

The case of algebraic numbers is of special interest and has a long history. First, we define the (naïve) height of an algebraic number.

Definition 6.3. Let θ be an irrational, real algebraic number of degree d and let $a_d X^d + \cdots + a_1 X + a_0$ denotes its minimal polynomial over $\mathbb{Z}$ (that is, the integer polynomial of lowest positive degree, with coprime coefficients and positive leading coefficient, which vanishes at θ). Then, the height $H(\theta)$ of θ is defined by

$$H(\theta) := \max\{|a_0|, |a_1|, \ldots, |a_d|\}.$$

We begin by a result of Liouville [51, 52] proved in 1844, and alluded to in Section 1.

Theorem 6.4. *Let θ be an irrational, real algebraic number of degree d and height at most H. Then,*

$$\left|\theta - \frac{p}{q}\right| \ge \frac{1}{d^2 H(1+|\theta|)^{d-1} q^d} \tag{2.18}$$

for all rational numbers p/q with $q \ge 1$.

Proof. Inequality (2.18) is true when $|\theta - p/q| \ge 1$. Let p/q be a rational number satisfying $|\theta - p/q| < 1$. Denoting by $P(X)$ the minimal defining polynomial of θ over $\mathbb{Z}$, we have $P(p/q) \neq 0$ and $|q^d P(p/q)| \ge 1$. By Rolle's Theorem, there exists a real number t lying between θ and p/q such that

$$|P(p/q)| = |P(\theta) - P(p/q)| = |\theta - p/q| \times |P'(t)|.$$

Since $|t - \theta| \le 1$ and

$$|P'(t)| \le d^2 H(1+|\theta|)^{d-1},$$

the combination of these inequalities gives the theorem. □

Thue [67] established in 1909 the first significant improvement of Liouville's result. There was subsequent progress by Siegel, Dyson and Gelfond, until Roth [61] proved in 1955 that, as far as approximation by rational numbers is concerned, the irrational, real algebraic numbers do behave like almost all real numbers.

Theorem 6.5. *For every $\varepsilon > 0$ and every irrational real algebraic number θ, there exist at most finitely many rational numbers p/q with $q \ge 1$ and*

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{q^{2+\varepsilon}}. \tag{2.19}$$

For a prime number ℓ and a non-zero rational number x, we set $|x|_\ell := \ell^{-u}$, where $u \in \mathbb{Z}$ is the exponent of ℓ in the prime decomposition of x. Furthermore, we set $|0|_\ell = 0$. The next theorem, proved by Ridout [59], extends Theorem 6.5.

Theorem 6.6. *Let S be a finite set of prime numbers. Let θ be a real algebraic number. Let ε be a positive real number. The inequality*

$$\prod_{\ell \in S} |pq|_\ell \cdot \min\left\{1, \left|\theta - \frac{p}{q}\right|\right\} < \frac{1}{q^{2+\varepsilon}}$$

has only finitely many solutions in non-zero integers p, q.

Theorem 6.5 is ineffective, in the sense that its proofs do not allow us to compute explicitly an integer q_0 such that (2.19) has no solution with q greater than q_0. Nevertheless, we are able to bound explicitly the *number* of primitive solutions (that is, of solutions in coprime integers p and q) to inequality (2.19). The first result in this direction was proved in 1955 by Davenport and Roth [38]; see the proof of Theorem 12.1 for a recent estimate.

The Schmidt Subspace Theorem [62, 63, 64] is a powerful multidimensional extension of the Roth Theorem, with many outstanding applications [20, 26, 69]. We quote below a version of it which is suitable for our purpose, but the reader should keep in mind that there are more general formulations.

Theorem 6.7. *Let $m \geq 2$ be an integer. Let S be a finite set of prime numbers. Let $L_{1,\infty}, \ldots, L_{m,\infty}$ be m linearly independent linear forms with real algebraic coefficients. For any prime ℓ in S, let $L_{1,\ell}, \ldots, L_{m,\ell}$ be m linearly independent linear forms with integer coefficients. Let ε be a positive real number. Then, there exist an integer T and proper subspaces $S_1, \ldots, S_T$ of $\mathbb{Q}^m$ such that all the solutions $\underline{x} = (x_1, \ldots, x_m)$ in $\mathbb{Z}^m$ to the inequality*

$$\prod_{\ell \in S} \prod_{i=1}^{m} |L_{i,\ell}(\underline{x})|_\ell \cdot \prod_{i=1}^{m} |L_{i,\infty}(\underline{x})| \leq (\max\{1, |x_1|, \ldots, |x_m|\})^{-\varepsilon} \tag{2.20}$$

are contained in the union $S_1 \cup \ldots \cup S_T$.

Let us briefly show how Roth's theorem can be deduced from Theorem 6.7. Let θ be a real algebraic number and ε be a positive real number. Consider the two independent linear forms $\theta X - Y$ and X. Theorem 6.7 implies that there are integers $T \geq 1, x_1, \ldots, x_T, y_1, \ldots, y_T$ with $(x_i, y_i) \neq (0, 0)$ for $i = 1, \ldots, T$, such that, for every integer solution (p, q) to

$$|q| \cdot |q\theta - p| < |q|^{-\varepsilon},$$

there exists an integer k with $1 \leq k \leq T$ and $x_k p + y_k q = 0$. If θ is irrational, this means that there are only finitely many rational solutions to $|\theta - p/q| < |q|^{-2-\varepsilon}$, which is Roth's theorem.

Note that (like Theorems 6.5 and 6.6) Theorem 6.7 is ineffective, in the sense that its proof does not yield an explicit upper bound for the height of the proper rational subspaces containing all the solutions to (2.20). Fortunately, Schmidt [65] was able to give an admissible value for the number T of subspaces; see [42] for a common generalization of Theorems 6.6 and 6.7, usually called the Quantitative Subspace Theorem, and [41] for the current state of the art.

7 Sketch of proof and historical comments

Before proving Theorems 3.1 and 3.2, we wish to highlight the main ideas and explain how weaker results can be deduced from various statements given in Section 6. We focus only on b-ary expansions.

The general idea goes as follows. Let us assume that (2.5) does not hold. Then, the sequence of digits of our real number satisfies a certain combinatorial property. And a suitable transcendence criterion prevents the sequence of digits of an irrational algebraic numbers to fulfill the same combinatorial property.

Let us see how transcendence results listed in Section 6 apply to get a combinatorial transcendence criterion. We introduce some more notation. Let W be a finite word. For a positive integer ℓ, we write W^ℓ for the word $W \dots W$ (ℓ times repeated concatenation of the word W) and W^∞ for the infinite word constructed by concatenation of infinitely many copies of W. More generally, for any positive real number x, we denote by W^x the word $W^{\lfloor x \rfloor} W'$, where W' is the prefix of W of length $\lceil (x - \lfloor x \rfloor)|W| \rceil$. Here, $\lceil \cdot \rceil$ denotes the upper integer part function. In particular, we can write

$$aabaaaabaaaa = (aabaa)^{12/5} = (aabaaaabaa)^{6/5} = (aabaa)^{\log 10}.$$

Let $\mathbf{a} = (a_\ell)_{\ell \ge 1}$ be a sequence of elements from $\mathcal{A}$. Let $w > 1$ be a real number. We say that $\mathbf{a}$ satisfies Condition $(\spadesuit)_w$ if $\mathbf{a}$ is not

ultimately periodic and if there exist two sequences of finite words $(Z_n)_{n \ge 1}$, and $(W_n)_{n \ge 1}$ such that:

(i) For every $n \ge 1$, the word $W_n Z_n^w$ is a prefix of the word $\mathbf{a}$.

(ii) The sequence $(|W_n|/|Z_n|)_{n \ge 1}$ is bounded from above.

(iii) The sequence $(|Z_n|)_{n \ge 1}$ is increasing.

We say that $\mathbf{a}$ satisfies Condition $(\spadesuit)_\infty$ if it satisfies Condition $(\spadesuit)_w$ for every $w > 1$.

Our first result is an application of Theorem 6.4 providing a combinatorial condition on the sequence b-ary expansion of a real number which ensures that this number is trancendental.

Theorem 7.1. *Let $b \ge 2$ be an integer. Let $\mathbf{a} = (a_\ell)_{\ell \ge 1}$ be a sequence of elements from $\{0, 1, \dots, b-1\}$. If $\mathbf{a}$ satisfies Condition $(\spadesuit)_\infty$, then the real number*

$$\xi := \sum_{\ell=1}^{+\infty} \frac{a_\ell}{b^\ell}$$

is transcendental.

Proof. Let $w > 1$ be a real number. By assumption, there exist two sequences of finite words, $(Z_n)_{n \ge 1}$, and $(W_n)_{n \ge 1}$, and an integer C, such that $|W_n| \le C|Z_n|$ and $W_n Z_n^w$ is a prefix of $\mathbf{a}$ for $n \ge 1$. Let $n \ge 1$ be an integer. In particular, ξ is very close to the rational number ξ_n whose b-ary expansion is the eventually periodic word $W_n Z_n^\infty$. A rapid calculation shows that there is an integer p_n such that

$$\xi_n = \frac{p_n}{b^{|W_n|}(b^{|Z_n|} - 1)}$$

and

$$
\begin{aligned}
|\xi - \xi_n| = \left| \xi - \frac{p_n}{b^{|W_n|}(b^{|Z_n|}-1)} \right| &\le \frac{1}{b^{|W_n|+w|Z_n|}} \\
&\le \left(\frac{1}{b^{|W_n|}(b^{|Z_n|}-1)} \right)^{(|W_n|+w|Z_n|)/(|W_n|+|Z_n|)}.
\end{aligned}
$$

Since $|W_n| \le C|Z_n|$, the quantity $(|W_n| + w|Z_n|)/(|W_n| + |Z_n|)$ is bounded from below by $(C + w)/(C + 1)$. Consequently, we get

$$
\left| \xi - \frac{p_n}{b^{|W_n|}(b^{|Z_n|}-1)} \right| \le \left(\frac{1}{b^{|W_n|}(b^{|Z_n|}-1)} \right)^{(C+w)/(C+1)}. \qquad (2.21)
$$

Let d be the integer part of $(C + w)/(C + 1)$. Since (2.21) holds for every $n \ge 1$, it follows from Theorem 6.4 that ξ cannot be algebraic of degree $\le d - 1$. Since w can be taken arbitrarily large, one deduces that ξ must be transcendental. □

It is apparent from the proof of Theorem 7.1 that, if instead of Liouville's theorem we use Roth's (Theorem 6.5) or, even better, Ridout's Theorem 6.6, then the assumptions of Theorem 7.1 can be substantially weakened. Indeed, Roth's theorem is sufficient to establish the transcendence of ξ as soon as the exponent $(C + w)/(C + 1)$ strictly exceeds 2, that is, if $w > 2 + C$. We explain below how Ridout's theorem yields a much better result.

Theorem 7.2. *Let $b \ge 2$ be an integer. Let $\mathbf{a} = (a_\ell)_{\ell \ge 1}$ be a sequence of elements from $\{0, 1, \ldots, b - 1\}$. Let $w > 2$ be a real number. If* **a** *satisfies Condition $(\spadesuit)_w$, then the real number*

$$
\xi := \sum_{\ell=1}^{+\infty} \frac{a_\ell}{b^\ell}
$$

is transcendental.

Proof. We keep the notation of the proof of Theorem 7.1, where it is shown that, for any $n \ge 1$, we have

$$
\left| \xi - \frac{p_n}{b^{|W_n|}(b^{|Z_n|}-1)} \right| \le \frac{1}{b^{|W_n|+w|Z_n|}},
$$

that is,

$$
\begin{aligned}
b^{-|W_n|} \left| \xi - \frac{p_n}{b^{|W_n|}(b^{|Z_n|}-1)} \right| &\le \frac{1}{b^{2|W_n|+w|Z_n|}} \\
&< \left(\frac{1}{b^{|W_n|}(b^{|Z_n|}-1)} \right)^{(2|W_n|+w|Z_n|)/(|W_n|+|Z_n|)}.
\end{aligned}
$$

Observe that, for $n \ge 1$,

$$
\frac{2|W_n| + w|Z_n|}{|W_n| + |Z_n|} = 2 + \frac{(w-2)|Z_n|}{|W_n| + |Z_n|} \ge 2 + \frac{w-2}{C+1}.
$$

Hence, if we take for S the set of prime divisors of b and set $\varepsilon := (w-2)/(C+1)$, we conclude that there are infinitely many rational numbers p/q such that

$$\prod_{\ell \in S} |pq|_\ell \cdot \min\left\{1, \left|\xi - \frac{p}{q}\right|\right\} < \frac{1}{q^{2+\varepsilon}}.$$

By Theorem 6.6, this shows that ξ is transcendental, since $w > 2$. □

We postpone to Section 9 the proof that the conclusion of Theorem 7.2 remains true under the weaker assumption $w > 1$ (the reader can easily check that this corresponds exactly to Theorem 4.1).

8 A combinatorial lemma

The purpose of this section is to establish a combinatorial lemma which allows us to deduce Theorems 3.1 and 3.2 from Theorems 4.1 and 4.2.

Lemma 8.1. *Let* $\mathbf{w} = w_1 w_2 \ldots$ *be an infinite word over a finite or an infinite alphabet* $\mathcal{A}$ *such that*

$$\liminf_{n \to +\infty} \frac{p(n, \mathbf{w}, \mathcal{A})}{n} < +\infty.$$

Then, the word $\mathbf{w}$ *satisfies Condition* (♠) *defined in Section 4.*

Proof. By assumption, there exist an integer $C \geq 2$ and an infinite set $\mathcal{N}$ of positive integers such that

$$p(n, \mathbf{w}, \mathcal{A}) \leq Cn, \quad \text{for every } n \text{ in } \mathcal{N}. \tag{2.22}$$

This implies, in particular, that $\mathbf{w}$ is written over a finite alphabet.

Let n be in $\mathcal{N}$. By (2.22) and the *Schubfachprinzip*, there exists (at least) one block X_n of length n having (at least) two occurrences in the prefix of length $(C+1)n$ of $\mathbf{w}$. Thus, there are words W_n, W_n', B_n and B_n' such that $|W_n| < |W_n'|$ and

$$w_1 \ldots w_{(C+1)n} = W_n X_n B_n = W_n' X_n B_n'.$$

If $|W_n X_n| \leq |W_n'|$, then define V_n by the equality $W_n X_n V_n = W_n'$. Observe that

$$w_1 \ldots w_{(C+1)n} = W_n X_n V_n X_n B_n' \tag{2.23}$$

and

$$\frac{|V_n| + |W_n|}{|X_n|} \leq C. \tag{2.24}$$

Set $U_n := X_n$.

If $|W_n'| < |W_n X_n|$, then, recalling that $|W_n| < |W_n'|$, we define X_n' by $W_n' = W_n X_n'$. Since $X_n B_n = X_n' X_n B_n'$ and $|X_n'| < |X_n|$, the word X_n' is a strict prefix of X_n and X_n is the concatenation of at least two copies of X_n' and a (possibly empty) prefix of X_n'. Let t_n be the largest positive integer such that X_n begins with $2t_n$ copies of X_n'. Observe that

$$2t_n|X_n'| + 2|X_n'| \geq |X_n' X_n|,$$

thus

$$n = |X_n| \leq (2t_n + 1)|X_n'| \leq 3t_n|X_n'|.$$

Consequently, $W_n({X_n'}^{t_n})^2$ is a prefix of $\mathbf{w}$ such that

$$|{X_n'}^{t_n}| \geq n/3$$

and

$$\frac{|W_n|}{|{X_n'}^{t_n}|} \leq \frac{3}{n} \cdot \left((C+1)n - 2|{X_n'}^{t_n}|\right) \leq 3C + 1. \tag{2.25}$$

Set $U_n := {X_n'}^{t_n}$ and let V_n be the empty word.

It then follows from (2.23), (2.24), and (2.25) that, for every n in the infinite set $\mathcal{N}$,

$$W_n U_n V_n U_n \quad \text{is a prefix of } \mathbf{w}$$

with

$$|W_n| + |V_n| \leq (3C+1)\,|U_n|.$$

This shows that $\mathbf{w}$ satisfies Condition (♠). □

We are now in position to deduce Theorems 3.1 and 3.2 from Theorems 4.1 and 4.2.

Proof of Theorem 3.1. Let $b \geq 2$ be an integer and ξ be an irrational real number. Assume that $p(n, \xi, b)$ does not tend to infinity with n. It then follows from Lemma 8.1 that the

infinite word composed of the digits of ξ written in base b satisfies Condition (♠). Consequently, Theorem 4.1 asserts that ξ cannot be algebraic. By contraposition, we get the theorem. □

Proof of Theorem 3.2. Let α be a real number not algebraic of degree at most two. Assume that $p(n, \alpha)$ does not tend to infinity with n. It then follows from Lemma 8.1 that the

infinite word composed of the partial quotients of ξ satisfies Condition (♠). Furthermore, $p(1, \alpha)$ is finite, thus the sequence of partial quotients of α is bounded, say by M. It then follows from Theorem 5.3 that $q_\ell \leq (M+1)^\ell$ for $\ell \geq 1$, hence the sequence $(q_\ell^{1/\ell})_{\ell \geq 1}$ is bounded. Consequently, all the hypotheses of Theorem 4.2 are satisfied, and one concludes that α cannot be algebraic of degree at least three. By contraposition, we get the theorem. □

9 Proof of Theorem 4.1

We present two proofs of Theorem 4.1, which was originally established in [10].

Throughout this section, we set $|U_n| = u_n$, $|V_n| = v_n$ and $|W_n| = w_n$, for $n \geq 1$. We assume that ξ is algebraic and we derive a contradiction by a suitable application of Theorem 6.7.

First proof. Let $n \geq 1$ be an integer. We observe that the real number ξ is quite close to the rational number ξ_n whose b-ary expansion is the infinite word $W_n(U_nV_n)^\infty$. Indeed, there exists an integer p_n such that

$$\xi_n = \frac{p_n}{b^{w_n}(b^{u_n+v_n} - 1)}$$

and

$$|\xi - \xi_n| \leq \frac{1}{b^{w_n+v_n+2u_n}},$$

since ξ and ξ_n have the same first $w_n + v_n + 2u_n$ digits in their b-ary expansion. Consequently, we have

$$|b^{w_n+u_n+v_n}\xi - b^{w_n}\xi - p_n| = |b^{w_n}(b^{u_n+v_n} - 1)\xi - p_n| \leq b^{-u_n}.$$

Consider the three linearly independent linear forms with real algebraic coefficients

$$\begin{aligned} L_{1,\infty}(X_1, X_2, X_3) &= X_1, \\ L_{2,\infty}(X_1, X_2, X_3) &= X_2, \\ L_{3,\infty}(X_1, X_2, X_3) &= \xi X_1 - \xi X_2 - X_3. \end{aligned}$$

Evaluating them on the integer points $\mathbf{x}_n := (b^{w_n+u_n+v_n}, b^{w_n}, p_n)$, we get that

$$\prod_{1 \leq j \leq 3} |L_{j,\infty}(\mathbf{x}_n)| \leq b^{2w_n+v_n}. \tag{2.26}$$

For any prime number ℓ dividing b, consider the three linearly independent linear forms with integer coefficients

$$\begin{aligned} L_{1,\ell}(X_1, X_2, X_3) &= X_1, \\ L_{2,\ell}(X_1, X_2, X_3) &= X_2, \\ L_{3,\ell}(X_1, X_2, X_3) &= X_3. \end{aligned}$$

We get that

$$\prod_{\ell | b} \prod_{1 \leq j \leq 3} |L_{j,\ell}(\mathbf{x}_n)|_\ell \leq b^{-2w_n-u_n-v_n}. \tag{2.27}$$

Since $\mathbf{a}$ satisfies Condition (♠), we have

$$\liminf_{n \to +\infty} \frac{u_n}{w_n + u_n + v_n} > 0.$$

It then follows from (2.26) and (2.27) that there exists $\varepsilon > 0$ such that

$$\prod_{1 \le j \le 3} |L_{j,\infty}(\mathbf{x}_n)| \cdot \prod_{\ell | b} \prod_{1 \le j \le 3} |L_{j,\ell}(\mathbf{x}_n)|_\ell \le b^{-u_n}$$
$$\le \max\{b^{w_n+u_n+v_n}, b^{w_n}, p_n\}^{-\varepsilon},$$

for every $n \ge 1$.

Now we infer from Theorem 6.7 that all the points $\mathbf{x}_n$ lie in a finite number of proper subspaces of $\mathbb{Q}^3$. Thus, there exist a non-zero integer triple (z_1, z_2, z_3) and an infinite set of distinct positive integers $\mathcal{N}_1$ such that

$$z_1 b^{w_n+u_n+v_n} + z_2 b^{w_n} + z_3 p_n = 0, \tag{2.28}$$

for any n in $\mathcal{N}_1$.

Dividing (2.28) by $b^{w_n+u_n+v_n}$, we get

$$z_1 + z_2 b^{-u_n-v_n} + z_3 \frac{p_n}{b^{w_n+u_n+v_n}} = 0. \tag{2.29}$$

Since u_n tends to infinity with n, the sequence $(p_n/b^{w_n+u_n+v_n})_{n \ge 1}$ tends to ξ. Letting n tend to infinity along $\mathcal{N}_1$, we then infer from (2.29) that either ξ is rational, or $z_1 = z_3 = 0$. In the latter case, z_2 must be zero, a contradiction. This shows that ξ cannot be algebraic. □

Second proof. Here, we follow an alternative approach presented in [2].

Let p_n and p'_n be the rational integers defined by

$$\sum_{\ell=1}^{w_n+v_n+2u_n} \frac{a_\ell}{b^\ell} = \frac{p_n}{b^{w_n+v_n+2u_n}} \quad \text{and} \quad \sum_{\ell=1}^{w_n+u_n} \frac{a_\ell}{b^\ell} = \frac{p'_n}{b^{w_n+u_n}}.$$

Observe that there exist integers f_n and f'_n such that

$$p_n = a_{w_n+v_n+2u_n} + a_{w_n+v_n+2u_n-1} b + \cdots + a_{w_n+v_n+u_n+1} b^{u_n-1} + f_n b^{u_n} \tag{2.30}$$

and

$$p'_n = a_{w_n+u_n} + a_{w_n+u_n-1} b + \cdots + a_{w_n+1} b^{u_n-1} + f'_n b^{u_n}. \tag{2.31}$$

Since, by assumption,

$$a_{w_n+u_n+v_n+j} = a_{w_n+j}, \quad \text{for } j = 1, \ldots, u_n,$$

it follows from (2.30) and (2.31) that $p_n - p'_n$ is divisible by an integer multiple of b^{u_n}. Thus, for any prime number ℓ dividing b, the ℓ-adic distance between p_n and p'_n is very small and we have

$$|p_n - p'_n|_\ell \le |b|_\ell^{u_n}.$$

Furthermore, it is clear that

$$|b^{w_n+u_n}\xi - p'_n| < 1 \quad \text{and} \quad |b^{w_n+v_n+2u_n}\xi - p_n| < 1.$$

Consider now the four linearly independent linear forms with real algebraic coefficients

$$\begin{aligned} L_{1,\infty}(X_1, X_2, X_3, X_4) &= X_1, \\ L_{2,\infty}(X_1, X_2, X_3, X_4) &= X_2, \\ L_{3,\infty}(X_1, X_2, X_3, X_4) &= \xi X_1 - X_3, \\ L_{4,\infty}(X_1, X_2, X_3, X_4) &= \xi X_2 - X_4. \end{aligned}$$

Evaluating them on the integer points $\mathbf{x}_n := (b^{w_n+v_n+2u_n}, b^{u_n+w_n}, p_n, p'_n)$, we get that

$$\prod_{1\le j\le 4} |L_{j,\infty}(\mathbf{x}_n)| \le b^{2w_n+v_n+3u_n}. \tag{2.32}$$

For any prime number p dividing b, we consider the four linearly independent linear forms with integer coefficients

$$\begin{aligned} L_{1,\ell}(X_1, X_2, X_3, X_4) &= X_1, \\ L_{2,\ell}(X_1, X_2, X_3, X_4) &= X_2, \\ L_{3,\ell}(X_1, X_2, X_3, X_4) &= X_3, \\ L_{4,\ell}(X_1, X_2, X_3, X_4) &= X_4 - X_3. \end{aligned}$$

We get that

$$\prod_{\ell | b} \prod_{1\le j\le 4} |L_{j,\ell}(\mathbf{x}_n)|_\ell \le b^{-(2w_n+v_n+3u_n)}\, b^{-u_n}. \tag{2.33}$$

Since **a** satisfies Condition (♠), we have

$$\liminf_{n\to+\infty} \frac{u_n}{w_n + 2u_n + v_n} > 0.$$

It then follows from (2.32) and (2.33) that there exists $\varepsilon > 0$ such that

$$\begin{aligned} \prod_{1\le j\le 4} |L_{j,\infty}(\mathbf{x}_n)| \cdot \prod_{\ell | b} \prod_{1\le j\le 4} |L_{j,\ell}(\mathbf{x}_n)|_\ell &\le b^{-u_n} \\ &\le \max\{b^{w_n+v_n+2u_n}, b^{u_n+w_n}, p_n, p'_n\}^{-\varepsilon}, \end{aligned}$$

for every $n \ge 1$.

We then infer from Theorem 6.7 that all the points $\mathbf{x}_n$ lie in a finite number of proper subspaces of $\mathbb{Q}^4$. Thus, there exist a non-zero integer quadruple (z_1, z_2, z_3, z_4) and an infinite set of distinct positive integers $\mathcal{N}_1$ such that

$$z_1 b^{w_n+v_n+2u_n} + z_2 b^{u_n+w_n} + z_3 p_n + z_4 p'_n = 0, \tag{2.34}$$

for any n in $\mathcal{N}_1$.

Dividing (2.34) by $b^{w_n+v_n+2u_n}$, we get

$$z_1 + z_2 b^{-u_n-v_n} + z_3 \frac{p_n}{b^{w_n+v_n+2u_n}} + z_4 b^{-u_n-v_n} \frac{p'_n}{b^{u_n+w_n}} = 0. \tag{2.35}$$

Recall that u_n tends to infinity with n. Thus, the sequences $(p_n/b^{w_n+v_n+2u_n})_{n\geq 1}$ and $(p'_n/b^{u_n+w_n})_{n\geq 1}$ tend to ξ as n tends to infinity. Letting n tend to infinity along $\mathcal{N}_1$, we infer from (2.35) that either ξ is rational, or $z_1 = z_3 = 0$. In the latter case, we obtain that ξ is rational. This is a contradiction, since the sequence $(a_\ell)_{\ell\geq 1}$ is not ultimately periodic. Consequently, ξ cannot be algebraic. □

Also, the Schmidt Subspace Theorem was applied similarly as in the first proof of Theorem 4.1 by Troi and Zannier [68] to establish the transcendence of the number $\sum_{m\in\mathcal{S}} 2^{-m}$, where $\mathcal{S}$ denotes the set of integers which can be represented as sums of distinct terms $2^k + 1$, where $k \geq 1$.

Moreover, in his short paper *Some suggestions for further research* published in 1984, Mahler [53] suggested explicitly to apply the Schmidt Subspace Theorem exactly as in the first proof of Theorem 4.1 given in Section 9 to investigate whether the middle third Cantor set contains irrational algebraic elements or not. More precisely, he wrote:

> *A possible approach to this question consists in the study of the non-homogeneous linear expressions*
>
> $$|3^{p_r+P_r}X - 3^{p_r}X - N_r|.$$
>
> *It may be that a p-adic form of Schmidt's theorem on the rational approximations of algebraic numbers* [10] *holds for such expressions.*

The reference [10] above is Schmidt's book [64].

We end this section by mentioning an application of Theorem 4.1. Adamczewski and Rampersad [12] proved that the binary expansion of an algebraic number contains infinitely many occurrences of 7/3-powers. They also established that the ternary expansion of an

algebraic number contains infinitely many occurrences of squares, or infinitely many occurrences of one of the blocks 010 or 02120.

10 Proof of Theorem 4.2

We reproduce the proof given in [29].

Throughout, the constants implied in $\ll$ depend only on α. Assume that the sequences $(U_n)_{n\geq 1}$, $(V_n)_{n\geq 1}$, and $(W_n)_{n\geq 1}$ occurring in the definition of Condition (♠) are fixed. For $n \geq 1$, set $u_n = |U_n|$, $v_n = |V_n|$, and $w_n = |W_n|$. We assume that the real number $\alpha := [0; a_1, a_2, \dots]$ is algebraic of degree at least three. Set $p_{-1} = q_0 = 1$ and $q_{-1} = p_0 = 0$.

We observe that α admits infinitely many good quadratic approximants obtained by truncating its continued fraction expansion and completing by periodicity. Precisely, for every positive integer n, we define the sequence $(b_k^{(n)})_{k \ge 1}$ by

$$\begin{aligned} b_h^{(n)} &= a_h \qquad && \text{for } 1 \le h \le w_n + u_n + v_n, \\ b_{w_n+h+j(u_n+v_n)}^{(n)} &= a_{w_n+h} \qquad && \text{for } 1 \le h \le u_n + v_n \text{ and } j \ge 0. \end{aligned}$$

The sequence $(b_k^{(n)})_{k \ge 1}$ is ultimately periodic, with preperiod W_n and with period $U_n V_n$. Set

$$\alpha_n = \left[0; b_1^{(n)}, b_2^{(n)}, \dots, b_k^{(n)}, \dots\right]$$

and note that, since the first $w_n + 2u_n + v_n$ partial quotients of α and of α_n are the same, it follows from Corollary 5.2 that

$$|\alpha - \alpha_n| \le q_{w_n+2u_n+v_n}^{-2}. \tag{2.36}$$

Furthermore, Lemma 5.8 asserts that α_n is root of the quadratic polynomial

$$\begin{aligned} P_n(X) := (&q_{w_n-1} q_{w_n+u_n+v_n} - q_{w_n} q_{w_n+u_n+v_n-1}) X^2 \\ &- (q_{w_n-1} p_{w_n+u_n+v_n} - q_{w_n} p_{w_n+u_n+v_n-1} \\ &+ p_{w_n-1} q_{w_n+u_n+v_n} - p_{w_n} q_{w_n+u_n+v_n-1}) X \\ &+ (p_{w_n-1} p_{w_n+u_n+v_n} - p_{w_n} p_{w_n+u_n+v_n-1}). \end{aligned}$$

By (2.10), we have

$$\begin{aligned} &|(q_{w_n-1} q_{w_n+u_n+v_n} - q_{w_n} q_{w_n+u_n+v_n-1}) \alpha - (q_{w_n-1} p_{w_n+u_n+v_n} - q_{w_n} p_{w_n+u_n+v_n-1})| \\ &\quad \le \ q_{w_n-1} |q_{w_n+u_n+v_n} \alpha - p_{w_n+u_n+v_n}| + q_{w_n} |q_{w_n+u_n+v_n-1} \alpha - p_{w_n+u_n+v_n-1}| \\ &\quad \le 2\, q_{w_n}\, q_{w_n+u_n+v_n}^{-1} \end{aligned} \tag{2.37}$$

and, likewise,

$$\begin{aligned} &|(q_{w_n-1} q_{w_n+u_n+v_n} - q_{w_n} q_{w_n+u_n+v_n-1}) \alpha - (p_{w_n-1} q_{w_n+u_n+v_n} - p_{w_n} q_{w_n+u_n+v_n-1})| \\ &\quad \le \ q_{w_n+u_n+v_n} |q_{w_n-1} \alpha - p_{w_n-1}| + q_{w_n+u_n+v_n-1} |q_{w_n} \alpha - p_{w_n}| \\ &\quad \le 2\, q_{w_n}^{-1}\, q_{w_n+u_n+v_n}. \end{aligned} \tag{2.38}$$

Using (2.36), (2.37), and (2.38), we then get

$$\begin{aligned} |P_n(\alpha)| &= |P_n(\alpha) - P_n(\alpha_n)| \\ &= |(q_{w_n-1} q_{w_n+u_n+v_n} - q_{w_n} q_{w_n+u_n+v_n-1})(\alpha - \alpha_n)(\alpha + \alpha_n) \\ &\quad - (q_{w_n-1} p_{w_n+u_n+v_n} - q_{w_n} p_{w_n+u_n+v_n-1} \\ &\quad + p_{w_n-1} q_{w_n+u_n+v_n} - p_{w_n} q_{w_n+u_n+v_n-1})(\alpha - \alpha_n)| \end{aligned}$$

$$
\begin{aligned}
&= |(q_{w_n-1}q_{w_n+u_n+v_n} - q_{w_n}q_{w_n+u_n+v_n-1})\alpha \\
&\quad - (q_{w_n-1}p_{w_n+u_n+v_n} - q_{w_n}p_{w_n+u_n+v_n-1}) \\
&\quad + (q_{w_n-1}q_{w_n+u_n+v_n} - q_{w_n}q_{w_n+u_n+v_n-1})\alpha \\
&\quad - (p_{w_n-1}q_{w_n+u_n+v_n} - p_{w_n}q_{w_n+u_n+v_n-1}) \\
&\quad + (q_{w_n-1}q_{w_n+u_n+v_n} - q_{w_n}q_{w_n+u_n+v_n-1})(\alpha_n - \alpha)| \cdot |\alpha - \alpha_n| \\
&\ll |\alpha - \alpha_n| \cdot \big(q_{w_n}q_{w_n+u_n+v_n}^{-1} + q_{w_n}^{-1}q_{w_n+u_n+v_n} + q_{w_n}q_{w_n+u_n+v_n}|\alpha - \alpha_n|\big) \\
&\ll |\alpha - \alpha_n| q_{w_n}^{-1}\, q_{w_n+u_n+v_n} \\
&\ll q_{w_n}^{-1}\, q_{w_n+u_n+v_n}\, q_{w_n+2u_n+v_n}^{-2}. \qquad (2.39)
\end{aligned}
$$

We consider the four linearly independent linear forms

$$
\begin{aligned}
L_1(X_1, X_2, X_3, X_4) &= \alpha^2 X_1 - \alpha(X_2 + X_3) + X_4, \\
L_2(X_1, X_2, X_3, X_4) &= \alpha X_1 - X_2, \\
L_3(X_1, X_2, X_3, X_4) &= \alpha X_1 - X_3, \\
L_4(X_1, X_2, X_3, X_4) &= X_1.
\end{aligned}
$$

Evaluating them on the 4-tuple

$$
\begin{aligned}
\mathbf{x}_n := (&q_{w_n-1}q_{w_n+u_n+v_n} - q_{w_n}q_{w_n+u_n+v_n-1}, \\
&q_{w_n-1}p_{w_n+u_n+v_n} - q_{w_n}p_{w_n+u_n+v_n-1}, \\
&p_{w_n-1}q_{w_n+u_n+v_n} - p_{w_n}q_{w_n+u_n+v_n-1}, \\
&p_{w_n-1}p_{w_n+u_n+v_n} - p_{w_n}p_{w_n+u_n+v_n-1}),
\end{aligned}
$$

it follows from (2.37), (2.38), (2.39), and Theorem 5.3 that

$$
\begin{aligned}
\prod_{1 \le j \le 4} |L_j(\mathbf{x}_n)| &\ll q_{w_n+u_n+v_n}^{2}\, q_{w_n+2u_n+v_n}^{-2} \\
&\ll 2^{-u_n} \\
&\ll (q_{w_n}q_{w_n+u_n+v_n})^{-\delta u_n/(2w_n+u_n+v_n)},
\end{aligned}
$$

if n is sufficiently large, where we have set

$$
M = 1 + \limsup_{\ell \to +\infty} q_\ell^{1/\ell} \quad \text{and} \quad \delta = \frac{\log 2}{\log M}.
$$

Since $\mathbf{a}$ satisfies Condition (♠), we have

$$
\liminf_{n \to +\infty} \frac{u_n}{2w_n + u_n + v_n} > 0.
$$

Consequently, there exists $\varepsilon > 0$ such that

$$
\prod_{1 \le j \le 4} |L_j(\mathbf{x}_n)| \ll (q_{w_n}\, q_{w_n+u_n+v_n})^{-\varepsilon}
$$

holds for any sufficiently large integer n.

It then follows from Theorem 6.7 that the points $\mathbf{x}_n$ lie in a finite union of proper linear subspaces of $\mathbb{Q}^4$. Thus, there exist a non-zero integer quadruple (x_1, x_2, x_3, x_4) and an infinite set $\mathcal{N}_1$ of distinct positive integers such that

$$\begin{aligned} & x_1(q_{w_n-1}q_{w_n+u_n+v_n} - q_{w_n}q_{w_n+u_n+v_n-1}) \\ & +x_2(q_{w_n-1}p_{w_n+u_n+v_n} - q_{w_n}p_{w_n+u_n+v_n-1}) \\ & +x_3(p_{w_n-1}q_{w_n+u_n+v_n} - p_{w_n}q_{w_n+u_n+v_n-1}) \\ & +x_4(p_{w_n-1}p_{w_n+u_n+v_n} - p_{w_n}p_{w_n+u_n+v_n-1}) = 0, \end{aligned} \tag{2.40}$$

for any n in $\mathcal{N}_1$.

- First case: we assume that there exist an integer ℓ and infinitely many integers n in $\mathcal{N}_1$ with $w_n = \ell$.

By extracting an infinite subset of $\mathcal{N}_1$ if necessary and by considering the real number $[0; a_{\ell+1}, a_{\ell+2}, \ldots]$ instead of α, we may without loss of generality assume that $w_n = \ell = 0$ for any n in $\mathcal{N}_1$.

Then, recalling that $q_{-1} = p_0 = 0$ and $q_0 = p_{-1} = 1$, we deduce from (2.40) that

$$x_1 q_{u_n+v_n-1} + x_2 p_{u_n+v_n-1} - x_3 q_{u_n+v_n} - x_4 p_{u_n+v_n} = 0, \tag{2.41}$$

for any n in $\mathcal{N}_1$. Observe that $(x_1, x_2) \neq (0, 0)$, since, otherwise, by letting n tend to infinity along $\mathcal{N}_1$ in (2.41), we would get that the real number α is rational. Dividing (2.41) by $q_{u_n+v_n}$, we obtain

$$x_1 \frac{q_{u_n+v_n-1}}{q_{u_n+v_n}} + x_2 \frac{p_{u_n+v_n-1}}{q_{u_n+v_n-1}} \cdot \frac{q_{u_n+v_n-1}}{q_{u_n+v_n}} - x_3 - x_4 \frac{p_{u_n+v_n}}{q_{u_n+v_n}} = 0. \tag{2.42}$$

By letting n tend to infinity along $\mathcal{N}_1$ in (2.42), we get that

$$\beta := \lim_{\mathcal{N}_1 \ni n \to +\infty} \frac{q_{u_n+v_n-1}}{q_{u_n+v_n}} = \frac{x_3 + x_4\alpha}{x_1 + x_2\alpha}.$$

Furthermore, observe that, for any sufficiently large integer n in $\mathcal{N}_1$, we have

$$\begin{aligned} \left| \beta - \frac{q_{u_n+v_n-1}}{q_{u_n+v_n}} \right| &= \left| \frac{x_3 + x_4\alpha}{x_1 + x_2\alpha} - \frac{x_3 + x_4 p_{u_n+v_n}/q_{u_n+v_n}}{x_1 + x_2 p_{u_n+v_n-1}/q_{u_n+v_n-1}} \right| \\ &\ll \frac{1}{q_{u_n+v_n-1}q_{u_n+v_n}}, \end{aligned} \tag{2.43}$$

by (2.10). Since the rational number $q_{u_n+v_n-1}/q_{u_n+v_n}$ is in its reduced form and $u_n + v_n$ tends to infinity when n tends to infinity along $\mathcal{N}_1$, we see that, for every positive real number η and every positive integer N, there exists a reduced rational number a/b such that $b > N$ and $|\beta - a/b| \le \eta/b$. This implies that β is irrational.

Consider now the three linearly independent linear forms

$$L'_1(Y_1, Y_2, Y_3) = \beta Y_1 - Y_2, \quad L'_2(Y_1, Y_2, Y_3) = \alpha Y_1 - Y_3, \quad L'_3(Y_1, Y_2, Y_3) = Y_2.$$

Evaluating them on the triple $(q_{u_n+v_n}, q_{u_n+v_n-1}, p_{u_n+v_n})$ with $n \in \mathcal{N}_1$, we infer from (2.10) and (2.43) that

$$\prod_{1 \le j \le 3} |L'_j(q_{u_n+v_n}, q_{u_n+v_n-1}, p_{u_n+v_n})| \ll q_{u_n+v_n}^{-1}.$$

It then follows from Theorem 6.7 that the points $(q_{u_n+v_n}, q_{u_n+v_n-1}, p_{u_n+v_n})$ with $n \in \mathcal{N}_1$ lie in a finite union of proper linear subspaces of $\mathbb{Q}^3$. Thus, there exist a non-zero integer triple (y_1, y_2, y_3) and an infinite set of distinct positive integers $\mathcal{N}_2 \subset \mathcal{N}_1$ such that

$$y_1 q_{u_n+v_n} + y_2 q_{u_n+v_n-1} + y_3 p_{u_n+v_n} = 0, \tag{2.44}$$

for any n in $\mathcal{N}_2$. Dividing (2.44) by $q_{u_n+v_n}$ and letting n tend to infinity along $\mathcal{N}_2$, we get

$$y_1 + y_2\beta + y_3\alpha = 0. \tag{2.45}$$

To obtain another equation relating α and β, we consider the three linearly independent linear forms

$$L''_1(Z_1, Z_2, Z_3) = \beta Z_1 - Z_2, \quad L''_2(Z_1, Z_2, Z_3) = \alpha Z_2 - Z_3,$$
$$L''_3(Z_1, Z_2, Z_3) = Z_2.$$

Evaluating them on the triple $(q_{u_n+v_n}, q_{u_n+v_n-1}, p_{u_n+v_n-1})$ with n in $\mathcal{N}_1$, we infer from (2.10) and (2.43) that

$$\prod_{1 \le j \le 3} |L''_j(q_{u_n+v_n}, q_{u_n+v_n-1}, p_{u_n+v_n-1})| \ll q_{u_n+v_n}^{-1}.$$

It then follows from Theorem 6.7 that the points $(q_{u_n+v_n}, q_{u_n+v_n-1}, p_{u_n+v_n-1})$ with $n \in \mathcal{N}_1$ lie in a finite union of proper linear subspaces of $\mathbb{Q}^3$. Thus, there exist a non-zero integer triple (z_1, z_2, z_3) and an infinite set of distinct positive integers $\mathcal{N}_3 \subset \mathcal{N}_2$ such that

$$z_1 q_{u_n+v_n} + z_2 q_{u_n+v_n-1} + z_3 p_{u_n+v_n-1} = 0, \tag{2.46}$$

for any n in $\mathcal{N}_3$. Dividing (2.46) by $q_{u_n+v_n-1}$ and letting n tend to infinity along $\mathcal{N}_3$, we get

$$\frac{z_1}{\beta} + z_2 + z_3\alpha = 0. \tag{2.47}$$

We infer from (2.45) and (2.47) that

$$(z_3\alpha + z_2)(y_3\alpha + y_1) = y_2 z_1.$$

Since β is irrational, we get from (2.45) and (2.47) that $y_3 z_3 \neq 0$. This shows that α is an algebraic number of degree at most two, which contradicts our assumption that α is algebraic of degree at least three.

- Second case: extracting an infinite subset $\mathcal{N}_4$ of $\mathcal{N}_1$ if necessary, we assume that $(w_n)_{n \in \mathcal{N}_4}$ tends to infinity.

In particular $(p_{w_n}/q_{w_n})_{n \in \mathcal{N}_4}$ and $(p_{w_n+u_n+v_n}/q_{w_n+u_n+v_n})_{n \in \mathcal{N}_4}$ both tend to α as n tends to infinity.

We make the following observation. Let a be a letter and U, V, W be three finite words (V may be empty) such that $\mathbf{a}$ begins with $WUVU$ and a is the last letter of W and of UV. Then, writing $W = W'a$, $V = V'a$ if V is non-empty, and $U = U'a$ if V is empty, we see that $\mathbf{a}$ begins with $W'(aU)V'(aU)$ if V is non-empty and with $W'(aU')(aU')$ if V is empty. Consequently, by iterating this remark if necessary, we can assume that for any n in $\mathcal{N}_4$, the last letter of the word $U_n V_n$ differs from the last letter of the word W_n. Said differently, we have $a_{w_n} \neq a_{w_n+u_n+v_n}$ for any n in $\mathcal{N}_4$.

Divide (2.40) by $q_{w_n} q_{w_n+u_n+v_n-1}$ and write

$$Q_n := (q_{w_n-1} q_{w_n+u_n+v_n})/(q_{w_n} q_{w_n+u_n+v_n-1}).$$

We then get

$$\begin{aligned} x_1(Q_n - 1) + x_2\left(Q_n \frac{p_{w_n+u_n+v_n}}{q_{w_n+u_n+v_n}} - \frac{p_{w_n+u_n+v_n-1}}{q_{w_n+u_n+v_n-1}}\right) & \\ + x_3\left(Q_n \frac{p_{w_n-1}}{q_{w_n-1}} - \frac{p_{w_n}}{q_{w_n}}\right) & \\ + x_4\left(Q_n \frac{p_{w_n-1}}{q_{w_n-1}} \frac{p_{w_n+u_n+v_n}}{q_{w_n+u_n+v_n}} - \frac{p_{w_n}}{q_{w_n}} \frac{p_{w_n+u_n+v_n-1}}{q_{w_n+u_n+v_n-1}}\right) &= 0, \end{aligned} \tag{2.48}$$

for any n in $\mathcal{N}_4$. To shorten the notation, for any $\ell \geq 1$, we put $R_\ell := \alpha - p_\ell/q_\ell$ and rewrite (2.48) as

$$\begin{aligned} x_1(Q_n - 1) + x_2\big(Q_n(\alpha - R_{w_n+u_n+v_n}) - (\alpha - R_{w_n+u_n+v_n-1})\big) \\ + x_3\big(Q_n(\alpha - R_{w_n-1}) - (\alpha - R_{w_n})\big) \\ + x_4\big(Q_n(\alpha - R_{w_n-1})(\alpha - R_{w_n+u_n+v_n}) - (\alpha - R_{w_n})(\alpha - R_{w_n+u_n+v_n-1})\big) = 0. \end{aligned}$$

This yields

$$\begin{aligned} (Q_n - 1)\big(x_1 + (x_2 + x_3)\alpha + x_4\alpha^2\big) = {} & x_2 Q_n R_{w_n+u_n+v_n} - x_2 R_{w_n+u_n+v_n-1} \\ & + x_3 Q_n R_{w_n-1} - x_3 R_{w_n} \\ & - x_4 Q_n R_{w_n-1} R_{w_n+u_n+v_n} \\ & + x_4 R_{w_n} R_{w_n+u_n+v_n-1} + \alpha(x_4 Q_n R_{w_n-1} \\ & + x_4 Q_n R_{w_n+u_n+v_n} - x_4 R_{w_n} \\ & - x_4 R_{w_n+u_n+v_n-1}). \end{aligned} \tag{2.49}$$

Observe that

$$|R_\ell| \leq q_\ell^{-1} q_{\ell+1}^{-1}, \quad \ell \geq 1, \tag{2.50}$$

by (2.10).

We use (2.49), (2.50) and the assumption that $a_{w_n} \neq a_{w_n+u_n+v_n}$ for any n in $\mathcal{N}_4$ to establish the following claim.

Claim. *We have*

$$x_1 + (x_2 + x_3)\alpha + x_4\alpha^2 = 0.$$

Proof of the Claim. If there are arbitrarily large integers n in $\mathcal{N}_4$ such that $Q_n \geq 2$ or $Q_n \leq 1/2$, then the claim follows from (2.49) and (2.50).

Assume that $1/2 \leq Q_n \leq 2$ holds for every large n in $\mathcal{N}_4$. We then derive from (2.49) and (2.50) that

$$|(Q_n - 1)(x_1 + (x_2 + x_3)\alpha + x_4\alpha^2)| \ll |R_{w_n-1}| \ll q_{w_n-1}^{-1} q_{w_n}^{-1}.$$

If $x_1 + (x_2 + x_3)\alpha + x_4\alpha^2 \neq 0$, then we get

$$|Q_n - 1| \ll q_{w_n-1}^{-1} q_{w_n}^{-1}. \tag{2.51}$$

On the other hand, observe that, by Theorem 5.4, the rational number Q_n is the quotient of the two continued fractions $[a_{w_n+u_n+v_n}; a_{w_n+u_n+v_n-1}, \ldots, a_1]$ and $[a_{w_n}; a_{w_n-1}, \ldots, a_1]$. Since $a_{w_n+u_n+v_n} \neq a_{w_n}$, we have either $a_{w_n+u_n+v_n} - a_{w_n} \geq 1$ or $a_{w_n} - a_{w_n+u_n+v_n} \geq 1$. In the former case, we see that

$$Q_n \geq \frac{a_{w_n+u_n+v_n}}{a_{w_n} + \cfrac{1}{1 + \cfrac{1}{a_{w_n-2} + 1}}} \geq \frac{a_{w_n} + 1}{a_{w_n} + \cfrac{a_{w_n-2} + 1}{a_{w_n-2} + 2}} \geq 1 + \frac{1}{(a_{w_n} + 1)(a_{w_n-2} + 2)}.$$

In the latter case, we have

$$\frac{1}{Q_n} \geq \frac{a_{w_n} + \cfrac{1}{a_{w_n-1}+1}}{a_{w_n+u_n+v_n}+1} \geq 1 + \frac{1}{(a_{w_n-1}+1)(a_{w_n+u_n+v_n}+1)} \geq 1 + \frac{1}{(a_{w_n-1}+1)a_{w_n}}.$$

Consequently, in any case,

$$|Q_n - 1| \gg a_{w_n}^{-1} \min\{a_{w_n-2}^{-1}, a_{w_n-1}^{-1}\} \gg a_{w_n}^{-1}\, q_{w_n-1}^{-1}.$$

Combined with (2.51), this gives

$$a_{w_n} \gg q_{w_n} \gg a_{w_n} q_{w_n-1},$$

which implies that n is bounded, a contradiction. This proves the Claim. □

Since α is irrational and not quadratic, we deduce from the Claim that $x_1 = x_4 = 0$ and $x_2 = -x_3$. Then, x_2 is non-zero and, by (2.40), we have, for any n in $\mathcal{N}_4$,

$$q_{w_n-1} p_{w_n+u_n+v_n} - q_{w_n} p_{w_n+u_n+v_n-1} = p_{w_n-1} q_{w_n+u_n+v_n} - p_{w_n} q_{w_n+u_n+v_n-1}.$$

Thus, the polynomial $P_n(X)$ can be simply expressed as

$$\begin{aligned}P_n(X) := (&q_{w_n-1}q_{w_n+u_n+v_n} - q_{w_n}q_{w_n+u_n+v_n-1})X^2\\ &- 2(q_{w_n-1}p_{w_n+u_n+v_n} - q_{w_n}p_{w_n+u_n+v_n-1})X\\ &+ (p_{w_n-1}p_{w_n+u_n+v_n} - p_{w_n}p_{w_n+u_n+v_n-1}).\end{aligned}$$

Consider now the three linearly independent linear forms

$$\begin{aligned}L_1'''(T_1, T_2, T_3) &= \alpha^2 T_1 - 2\alpha T_2 + T_3,\\ L_2'''(T_1, T_2, T_3) &= \alpha T_1 - T_2,\\ L_3'''(T_1, T_2, T_3) &= T_1.\end{aligned}$$

Evaluating them on the triple

$$\begin{aligned}\mathbf{x}_n' := (&q_{w_n-1}q_{w_n+u_n+v_n} - q_{w_n}q_{w_n+u_n+v_n-1},\\ &q_{w_n-1}p_{w_n+u_n+v_n} - q_{w_n}p_{w_n+u_n+v_n-1},\\ &p_{w_n-1}p_{w_n+u_n+v_n} - p_{w_n}p_{w_n+u_n+v_n-1}),\end{aligned}$$

for n in $\mathcal{N}_4$, it follows from (2.37) and (2.39) that

$$\prod_{1\le j\le 3} |L_j'''(\mathbf{x}_n')| \ll q_{w_n}\, q_{w_n+u_n+v_n}\, q_{w_n+2u_n+v_n}^{-2} \ll (q_{w_n}\, q_{w_n+u_n+v_n})^{-\varepsilon},$$

with the same ε as above, if n is sufficiently large.

We then deduce from Theorem 6.7 that the points $\mathbf{x}_n'$, $n \in \mathcal{N}_4$, lie in a finite union of proper linear subspaces of $\mathbb{Q}^3$. Thus, there exist a non-zero integer triple (t_1, t_2, t_3) and an infinite set of distinct positive integers $\mathcal{N}_5$ included in $\mathcal{N}_4$ such that

$$\begin{aligned}&t_1(q_{w_n-1}q_{w_n+u_n+v_n} - q_{w_n}q_{w_n+u_n+v_n-1})\\ &+t_2(q_{w_n-1}p_{w_n+u_n+v_n} - q_{w_n}p_{w_n+u_n+v_n-1})\\ &+t_3(p_{w_n-1}p_{w_n+u_n+v_n} - p_{w_n}p_{w_n+u_n+v_n-1}) = 0,\end{aligned}\tag{2.52}$$

for any n in $\mathcal{N}_5$.

We proceed exactly as above. Divide (2.52) by $q_{w_n}\, q_{w_n+u_n+v_n-1}$ and set

$$Q_n := (q_{w_n-1}q_{w_n+u_n+v_n})/(q_{w_n}q_{w_n+u_n+v_n-1}).$$

We then get

$$\begin{aligned}t_1(Q_n - 1) + t_2\biggl(Q_n \frac{p_{w_n+u_n+v_n}}{q_{w_n+u_n+v_n}} - \frac{p_{w_n+u_n+v_n-1}}{q_{w_n+u_n+v_n-1}}\biggr)&\\ + t_3\biggl(Q_n \frac{p_{w_n-1}}{q_{w_n-1}}\frac{p_{w_n+u_n+v_n}}{q_{w_n+u_n+v_n}} - \frac{p_{w_n}}{q_{w_n}}\frac{p_{w_n+u_n+v_n-1}}{q_{w_n+u_n+v_n-1}}\biggr) &= 0,\end{aligned}\tag{2.53}$$

for any n in $\mathcal{N}_5$. We argue as after (2.48). Since p_{w_n}/q_{w_n} and $p_{w_n+u_n+v_n}/q_{w_n+u_n+v_n}$ tend to α as n tends to infinity along $\mathcal{N}_5$, we derive from (2.53) that

$$t_1 + t_2\alpha + t_3\alpha^2 = 0,$$

a contradiction since α is irrational and not quadratic. Consequently, α must be transcendental. This concludes the proof of the theorem. □

11 A transcendence criterion for quasi-palindromic continued fractions

In this section, we present another combinatorial transcendence criterion for continued fractions which was established in [29], based on ideas from [5].

For a finite word $W := w_1 \dots w_k$, we denote by $\overline{W} := w_k \dots w_1$ its mirror image. The finite word W is called a *palindrome* if $W = \overline{W}$.

Let $\mathbf{a} = (a_\ell)_{\ell \geq 1}$ be a sequence of elements from $\mathcal{A}$. We say that $\mathbf{a}$ satisfies Condition (♣) if $\mathbf{a}$ is not

ultimately periodic and if there exist three sequences of finite words $(U_n)_{n \geq 1}$, $(V_n)_{n \geq 1}$, and $(W_n)_{n \geq 1}$ such that:

(i) For every $n \geq 1$, the word $W_n U_n V_n \overline{U}_n$ is a prefix of the word $\mathbf{a}$.

(ii) The sequence $(|V_n|/|U_n|)_{n \geq 1}$ is bounded from above.

(iii) The sequence $(|W_n|/|U_n|)_{n \geq 1}$ is bounded from above.

(iv) The sequence $(|U_n|)_{n \geq 1}$ is increasing.

Theorem 11.1. *Let $\mathbf{a} = (a_\ell)_{\ell \geq 1}$ be a sequence of positive integers. Let $(p_\ell/q_\ell)_{\ell \geq 1}$ denote the sequence of convergents to the real number*

$$\alpha := [0; a_1, a_2, \dots, a_\ell, \dots].$$

Assume that the sequence $(q_\ell^{1/\ell})_{\ell \geq 1}$ is bounded. If $\mathbf{a}$ satisfies Condition (♣)*, then α is transcendental.*

A slight modification of the proof of Theorem 11.1 allows us to remove the assumption on the growth of the sequence $(q_\ell)_{\ell \geq 1}$, provided that a stronger condition than Condition (♣) is satisfied.

Theorem 11.2. *Let $\mathbf{a} = (a_\ell)_{\ell \geq 1}$ be a sequence of positive integers and set*

$$\alpha := [0; a_1, a_2, \dots, a_\ell, \dots].$$

Assume that $\mathbf{a} = (a_\ell)_{\ell \geq 1}$ is not eventually periodic. If there are arbitrarily large integers ℓ such that the word $a_1 \dots a_\ell$ is a palindrome, then α is transcendental.

We leave to the reader the proof of Theorem 11.2, established in [5], and establish Theorem 11.1.

Proof. Throughout, the constants implied in $\ll$ are absolute.

Assume that the sequences $(U_n)_{n \geq 1}$, $(V_n)_{n \geq 1}$, and $(W_n)_{n \geq 1}$ are fixed. Set $w_n = |W_n|$, $u_n = |U_n|$ and $v_n = |V_n|$, for $n \geq 1$. Assume that the real number $\alpha := [0; a_1, a_2, \dots]$ is algebraic of degree at least three.

For $n \geq 1$, consider the rational number P_n/Q_n defined by

$$\frac{P_n}{Q_n} := [0; W_n U_n V_n \overline{U_n}\, \overline{W_n}]$$

and denote by P'_n/Q'_n the last convergent to P_n/Q_n which is different from P_n/Q_n. Since the first $w_n + 2u_n + v_n$ partial quotients of α and P_n/Q_n coincide, we deduce from Corollary 5.2 that

$$|Q_n\alpha - P_n| < Q_n q^{-2}_{w_n+2u_n+v_n}, \quad |Q'_n\alpha - P'_n| < Q_n q^{-2}_{w_n+2u_n+v_n}. \tag{2.54}$$

Furthermore, it follows from Theorem 5.4 that the first $w_n + 2u_n + v_n$
partial quotients of α and of Q'_n/Q_n coincide, thus

$$|Q_n\alpha - Q'_n| < Q_n q^{-2}_{w_n+u_n}, \tag{2.55}$$

again by Corollary 5.2, and, by Theorem 5.6,

$$Q_n \le 2q_{w_n}q_{w_n+2u_n+v_n} \le 2q_{w_n+u_n}q_{w_n+2u_n+v_n}. \tag{2.56}$$

Since

$$\begin{aligned}\alpha(Q_n\alpha - P_n) - (Q'_n\alpha - P'_n) &= \alpha Q_n\left(\alpha - \frac{P_n}{Q_n}\right) - Q'_n\left(\alpha - \frac{P'_n}{Q'_n}\right)\\ &= (\alpha Q_n - Q'_n)\left(\alpha - \frac{P_n}{Q_n}\right) + Q'_n\left(\frac{P'_n}{Q'_n} - \frac{P_n}{Q_n}\right),\end{aligned}$$

it follows from (2.54), (2.55), and (2.56) that

$$\begin{aligned}|\alpha^2 Q_n - \alpha Q'_n - \alpha P_n + P'_n| &\ll Q_n q^{-2}_{w_n+u_n} q^{-2}_{w_n+2u_n+v_n} + Q_n^{-1}\\ &\ll Q_n^{-1}.\end{aligned} \tag{2.57}$$

Consider the four linearly independent linear forms with algebraic coefficients

$$\begin{aligned}L_1(X_1, X_2, X_3, X_4) &= \alpha^2 X_1 - \alpha X_2 - \alpha X_3 + X_4,\\ L_2(X_1, X_2, X_3, X_4) &= \alpha X_2 - X_4,\\ L_3(X_1, X_2, X_3, X_4) &= \alpha X_1 - X_2,\\ L_4(X_1, X_2, X_3, X_4) &= X_2.\end{aligned}$$

We deduce from (2.54), (2.55), (2.56), and (2.57) that

$$\prod_{1\le j\le 4} |L_j(Q_n, Q'_n, P_n, P'_n)| \ll Q_n^2\, q^{-2}_{w_n+2u_n+v_n}\, q^{-2}_{w_n+u_n} \ll q^2_{w_n}\, q^{-2}_{w_n+u_n}.$$

By combining Theorems 5.3 and 5.6 with (2.56), we have

$$q^2_{w_n}\, q^{-2}_{w_n+u_n} \ll 2^{-u_n} \ll Q_n^{-\delta u_n/(2w_n+2u_n+v_n)},$$

if n is sufficiently large, where we have set

$$M = 1 + \limsup_{\ell\to+\infty} q_\ell^{1/\ell} \quad \text{and} \quad \delta = \frac{\log 2}{\log M}.$$

Since **a** satisfies Condition (♣), we have

$$\liminf_{n\to+\infty} \frac{u_n}{2w_n + 2u_n + v_n} > 0.$$

Consequently, there exists $\varepsilon > 0$ such that

$$\prod_{1\le j\le 4} |L_j(Q_n, Q'_n, P_n, P'_n)| \ll Q_n^{-\varepsilon},$$

for every sufficiently large n.

It then follows from Theorem 6.7 that the points (Q_n, Q'_n, P_n, P'_n) lie in a finite number of proper subspaces of $\mathbb{Q}^4$. Thus, there exist a non-zero integer 4-tuple (x_1, x_2, x_3, x_4) and an infinite set of distinct positive integers $\mathcal{N}_1$ such that

$$x_1 Q_n + x_2 Q'_n + x_3 P_n + x_4 P'_n = 0, \tag{2.58}$$

for any n in $\mathcal{N}_1$. Dividing by Q_n, we obtain

$$x_1 + x_2 \frac{Q'_n}{Q_n} + x_3 \frac{P_n}{Q_n} + x_4 \frac{P'_n}{Q'_n} \cdot \frac{Q'_n}{Q_n} = 0.$$

By letting n tend to infinity along $\mathcal{N}_1$, we infer from (2.55) and (2.56) that

$$x_1 + (x_2 + x_3)\alpha + x_4\alpha^2 = 0.$$

Since $(x_1, x_2, x_3, x_4) \neq (0, 0, 0, 0)$ and since α is irrational and not quadratic, we have $x_1 = x_4 = 0$ and $x_2 = -x_3$. Then, (2.58) implies that

$$Q'_n = P_n.$$

for every n in $\mathcal{N}_1$. Thus, for n in $\mathcal{N}_1$, we have

$$|\alpha^2 Q_n - 2\alpha Q'_n + P'_n| \ll Q_n^{-1}. \tag{2.59}$$

Consider now the three linearly independent linear forms

$$\begin{aligned}
L'_1(X_1, X_2, X_3) &= \alpha^2 X_1 - 2\alpha X_2 + X_3,\\
L'_2(X_1, X_2, X_3) &= \alpha X_2 - X_3,\\
L'_3(X_1, X_2, X_3) &= X_1.
\end{aligned}$$

Evaluating them on the triple (Q_n, Q'_n, P'_n) for n in $\mathcal{N}_1$, it follows from (2.54), (2.56) and (2.59) that

$$\begin{aligned}
\prod_{1\le j\le 3} |L'_j(Q_n, Q'_n, P'_n)| &\ll Q_n q_{w_n+2u_n+v_n}^{-2}\\
&\ll q_{w_n} q_{w_n+2u_n+v_n}^{-1} \ll q_{w_n} q_{w_n+u_n}^{-1} \ll Q_n^{-\varepsilon/2},
\end{aligned}$$

with the same ε as above, if n is sufficiently large.

It then follows from Theorem 6.7 that the points (Q_n, Q'_n, P'_n) lie in a finite number of proper subspaces of $\mathbb{Q}^3$. Thus, there exist a non-zero integer triple (y_1, y_2, y_3) and an infinite set of distinct positive integers $\mathcal{N}_2$ such that

$$y_1 Q_n + y_2 Q'_n + y_3 P'_n = 0, \tag{2.60}$$

for any n in $\mathcal{N}_2$.

Dividing (2.60) by Q_n, we get

$$y_1 + y_2 \frac{P_n}{Q_n} + y_3 \frac{P'_n}{Q'_n} \cdot \frac{P_n}{Q_n} = 0. \tag{2.61}$$

By letting n tend to infinity along $\mathcal{N}_2$, it thus follows from (2.61) that

$$y_1 + y_2\alpha + y_3\alpha^2 = 0.$$

Since (y_1, y_2, y_3) is a non-zero triple of integers, we have reached a contradiction. Consequently, the real number α is transcendental. This completes the proof of the theorem. □

12 Complements

We collect in this section several results which complement Theorems 3.1 and 3.2. The common tool for their proofs (which we omit or just sketch) is the Quantitative Subspace Theorem, that is, a theorem which provides an explicit upper bound for the number T of exceptional subspaces in Theorem 6.7.

We have mentioned at the beginning of Section 3 that the sequence of partial quotients of an algebraic irrational number θ cannot grow too rapidly. More precisely, it can be derived from Roth's Theorem 6.5 that

$$\lim_{n\to+\infty} \frac{\log\log q_n}{n} = 0, \tag{2.62}$$

where $(p_\ell/q_\ell)_{\ell\geq 1}$ denotes the sequence of convergents to θ. This is left as an exercise. The use of a quantitative form of Theorem 6.5 allowed Davenport and Roth [38] to improve (2.62). Their result was subsequently strengthen [4, 25] as follows.

Theorem 12.1. *Let θ be an irrational, real algebraic number and let $(p_n/q_n)_{n\geq 1}$ denote the sequence of its convergents. Then, for any $\varepsilon > 0$, there exists a constant c, depending only on θ and ε, such that*

$$\log\log q_n \leq c\, n^{2/3+\varepsilon}.$$

Proof. We briefly sketch the proof of a slightly weaker result. Let d be the degree of θ. By Theorem 6.4, there exists an integer n_0 such that

$$\left|\theta - \frac{p_n}{q_n}\right| > \frac{1}{q_n^{d+1}},$$

for $n \ge n_0$. Combined with Theorem 5.1, this gives

$$q_{n+1} \le q_n^d, \quad \text{for } n \ge n_0. \tag{2.63}$$

On the other hand, a quantitative form of Theorem 6.5 (see e.g. [40]) asserts that there exists a positive number $\eta_0 < 1/5$, depending only on θ, such that for every η with $0 < \eta < \eta_0$, the inequality

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{q^{2+\eta}},$$

has at most η^{-4} rational solutions p/q with p and q co prime and $q > 16^{1/\eta}$.

Consequently, for every $n \ge 8/\eta$, with at most η^{-4} exceptions, we have

$$q_{n+1} \le q_n^{1+\eta}. \tag{2.64}$$

Let $N \ge (8/\eta)^2$ be a large integer and set $h := \lceil \sqrt{N} \rceil$. We deduce from (2.63) and (2.64) that

$$\begin{aligned}\frac{\log q_N}{\log q_h} &= \frac{\log q_N}{\log q_{N-1}} \times \frac{\log q_{N-1}}{\log q_{N-2}} \times \cdots \times \frac{\log q_{h+1}}{\log q_h} \\ &\le (1+\eta)^N \, d^{\eta^{-4}},\end{aligned}$$

whence

$$\log\log q_N - \log\log q_h \le N\,\eta + \eta^{-4}(\log d).$$

Observe that it follows from (2.62) that

$$\log\log q_h \le N^{1/2},$$

when N is sufficiently large. Choosing $\eta = N^{-1/5}$, we obtain the upper bound

$$\log\log q_N \le \log\log q_h + N^{4/5}(\log 3d) \le 2N^{4/5}(\log 3d),$$

when N is large enough. A slight refinement yields the theorem. □

We first observe that, if we assume a slightly stronger condition than

$$\liminf_{n \to +\infty} \frac{p(n, \mathbf{w}, \mathcal{A})}{n} < +\infty$$

in Lemma 8.1, namely, that

$$\limsup_{n \to +\infty} \frac{p(n, \mathbf{w}, \mathcal{A})}{n} < +\infty,$$

then the word $\mathbf{w}$ satisfies a much stronger condition than Condition (♠). Indeed, there then exists an integer $C \ge 2$ such that

$$p(n, \mathbf{w}, \mathcal{A}) \le Cn, \quad \text{for } n \ge 1,$$

instead of the weaker assumption (2.22). In the case of the first proof of Theorem 4.1 given in Section 9, this means that, keeping its notation, one may assume that, up to extracting subsequences, there exists an integer c such that

$$2(u_n + v_n + w_n) \le u_{n+1} + v_{n+1} + w_{n+1} \le c(u_n + v_n + w_n), \quad n \ge 1.$$

This observation is crucial for the proofs of Theorems 12.2 to 12.4 below.

Now, we mention a few applications to the complexity of algebraic numbers, beginning with a result from [31]. Recall that the complexity function $n \mapsto p(n, \theta, b)$ has been defined in Section 2.

Theorem 12.2. *Let $b \ge 2$ be an integer and θ an algebraic irrational number. Then, for any real number η such that $\eta < 1/11$, we have*

$$\limsup_{n \to +\infty} \frac{p(n, \theta, b)}{n(\log n)^\eta} = +\infty.$$

The main tools for the proof of Theorem 12.2 are a suitable extension of the Cugiani–Mahler Theorem and a suitable version of the Quantitative Subspace Theorem, which allows us to get an exponent of $\log n$ independent of the base b. Using the recent results of [41] allows us to show that Theorem 12.2 holds for η in a slightly larger interval than $[0, 1/11)$.

As briefly mentioned in Section 6, one of the main features of the theorems of Roth and Schmidt is that they are ineffective, in the sense that we cannot produce an explicit upper bound for the denominators of the solutions to (2.19) or for the height of the subspaces containing the solutions to (2.20). Consequently, Theorems 3.1 and 3.2 are ineffective, as are the weaker results from [14, 43]. It is shown in [24] that, by means of the Quantitative Subspace Theorem, it is possible to derive an explicit form of a much weaker statement.

Theorem 12.3. *Let $b \ge 2$ be an integer. Let θ be a real algebraic irrational number of degree d and height at most H, with $H \ge e^e$. Set*

$$M = \exp\{10^{190} (\log(8d))^2 (\log\log(8d))^2\} + 2^{32 \log(240 \log(4H))}.$$

Then we have

$$p(n, \theta, b) \ge \left(1 + \frac{1}{M}\right) n, \quad \textit{for } n \ge 1.$$

Unfortunately, the present methods do not seem to be powerful enough to get an effective version of Theorem 3.1.

We have shown that, if the b-ary or the continued fraction expansion of a real number is not ultimately periodic and has small complexity, then this number cannot be algebraic, that is, the distance between this number and the set of algebraic numbers is strictly positive. A natural question then arises: is it possible to get transcendence measures for ξ, that is, to bound from below the distance between ξ and any algebraic number? A positive answer was given in [6], where the authors described a general method to obtain transcendence measures by means of the Quantitative Subspace Theorem. In the next statement, proved in [9], we say that an infinite word **w** written on an alphabet $\mathcal{A}$ is of sublinear complexity if there exists a constant C such that the complexity function of **w** satisfies

$$p(n, \mathbf{w}, \mathcal{A}) \leq Cn, \quad \text{for all } n \geq 1.$$

Recall that a Liouville number is an irrational real number γ such that for every real number w, there exists a rational number p/q with $|\gamma - p/q| < 1/q^w$.

Theorem 12.4. *Let ξ be an irrational real number and $b \geq 2$ be an integer. If the b-ary expansion of ξ is of sublinear complexity, then, either ξ is a Liouville number, or there exists a positive number C such that*

$$|\xi - \theta| > H(\theta)^{-(2d)^{C \log\log(3d)}},$$

for every real algebraic number θ of degree d.

In Theorem 12.4, the quantity $H(\theta)$ is the height of θ as introduced in Definition 6.3.

The analogue of Theorem 12.4 for continued fraction expansions has been established in [7, 28]. The analogues of Theorems 12.2 and 12.3 have not been written yet, but there is little doubt that they hold and can be proved by combining the ideas of the proofs of Theorems 3.2, 12.2 and 12.3.

13 Further notions of complexity

For an integer $b \geq 2$, an irrational real number ξ whose b-ary expansion is given by (2.1), and a positive integer n, set

$$\mathcal{NZ}(n, \xi, b) := \#\{\ell : 1 \leq \ell \leq n, a_\ell \neq 0\},$$

which counts the number of non-zero digits among the first n digits of the b-ary expansion of ξ.

Alternatively, if $1 \leq n_1 < n_2 < \cdots$ denotes the increasing sequence of the indices ℓ such that $a_\ell \neq 0$, then for every positive integer n we have

$$\mathcal{NZ}(n, \xi, b) := \max\{j : n_j \leq n\}.$$

Let $\varepsilon > 0$ be a real number and θ be an algebraic, irrational number. It follows from Ridout's Theorem 6.6 that $n_{j+1} \leq (1+\varepsilon)n_j$ holds for every sufficiently large j. Consequently, we get that

$$\lim_{n \to +\infty} \frac{\mathcal{NZ}(n, \theta, b)}{\log n} = +\infty.$$

For the base $b = 2$, this was considerably improved by Bailey, Borwein, Crandall, and Pomerance [17] (see also Rivoal [60]), using elementary considerations and ideas from additive number theory. A minor modification of their method allows us to get a similar result for expansions to an arbitrary integer base. The following statement is extracted from [27] (see also [11]).

Theorem 13.1. *Let $b \geq 2$ be an integer. For any irrational real algebraic number θ of degree d and height H and for any integer n exceeding $(20b^d d^3 H)^d$, we have*

$$\mathcal{NZ}(n, \theta, b) \geq \frac{1}{b-1}\left(\frac{n}{2(d+1)a_d}\right)^{1/d},$$

where a_d denotes the leading coefficient of the minimal polynomial of θ over the integers.

The idea behind the proof of Theorem 13.1 is quite simple and was inspired by a paper by Knight [48]. If an irrational real number ξ has few non-zero digits, then its integer powers $\xi^2, \xi^3, \ldots$, and any finite linear combination of them, cannot have too many non-zero digits. In particular, ξ cannot be a root of an integer polynomial of small degree. This is, in general, not at all so simple, since we have to take much care of the carries. However, for some particular families of algebraic numbers, including roots of positive integers, a quite simple proof of Theorem 13.1 can be given. Here, we follow [60] and (this allows some minor simplification) we treat only the case $b = 2$.

For a non-negative integer x, let $B(x)$ denote the number of 1's in the (finite) binary representation of x.

Theorem 13.2. *Let θ be a positive real algebraic number of degree $d \geq 2$. Let $a_d X^d + \cdots + a_1 X + a_0$ denote its minimal polynomial and assume that $a_1, \ldots, a_d$ are all non-negative. Then, there exists a constant c, depending only on θ, such that*

$$\mathcal{NZ}(n, \theta, 2) \geq B(a_d)^{-1/d} n^{1/d} - c,$$

for $n \geq 1$.

Proof. Observe first that, for all positive integers x and y, we have

$$B(x+y) \leq B(x) + B(y)$$

and

$$B(xy) \leq B(x)\, B(y).$$

For simplicity, let us write $\mathcal{NZ}(n, \cdot)$ instead of $\mathcal{NZ}(n, \cdot, 2)$. Let ξ and η be positive irrational numbers (the assumption of positivity is crucial) and n be a sufficiently large integer. We state without proof several elementary assertions. If $\xi + \eta$ is irrational, then we have

$$\mathcal{NZ}(n, \xi + \eta) \leq \mathcal{NZ}(n, \xi) + \mathcal{NZ}(n, \eta) + 1.$$

If $\xi\eta$ is irrational, then we have

$$\mathcal{NZ}(n,\xi\eta) \le \mathcal{NZ}(n,\xi)\cdot\mathcal{NZ}(n,\eta) + \log_2(\xi+\eta+1)+1,$$

where $\log_2$ denotes the logarithm in base 2. If m is an integer, then we have

$$\mathcal{NZ}(n,m\xi) \le B(m)(\mathcal{NZ}(n,\xi)+1).$$

Furthermore, for every positive integer A, we have

$$\mathcal{NZ}(n,\xi)\cdot\mathcal{NZ}(n,A/\xi) \ge n-1-\log_2(\xi+A/\xi+1)). \tag{2.65}$$

Let θ be as in the statement of the theorem. The real number $|a_0|/\theta$ is irrational, as are the numbers $a_j\theta^{j-1}$ for $j=2,\dots,d$ provided that $a_j \ne 0$. Since

$$|a_0|\theta^{-1} = a_1 + a_2\theta + \cdots + a_d\theta^{d-1}$$

and $\mathcal{NZ}(n,\theta)$ tends to infinity with n, the various inequalities listed above imply that

$$\begin{aligned}\mathcal{NZ}(n,|a_0|\theta^{-1}) &\le d + B(a_1) + \mathcal{NZ}(n,a_2\theta)+\cdots+\mathcal{NZ}(n,a_d\theta^{d-1})\\ &\le d + B(a_1) + B(a_2)(\mathcal{NZ}(n,\theta)+1)+\cdots\\ &\quad + B(a_d)(\mathcal{NZ}(n,\theta^{d-1})+1)\\ &\le B(a_d)\mathcal{NZ}(n,\theta)^{d-1} + c_1\mathcal{NZ}(n,\theta)^{d-2},\end{aligned} \tag{2.66}$$

where c_1, like c_2, c_3, c_4 below, is a suitable positive real number depending only on θ. By (2.65), we get

$$\mathcal{NZ}(n,|a_0|\theta^{-1}) \ge \frac{n}{\mathcal{NZ}(n,\theta)} - c_2.$$

Combining this with (2.66), we obtain

$$B(a_d)\mathcal{NZ}(n,\theta)^d + c_3\mathcal{NZ}(n,\theta)^{d-1} \ge n$$

and we finally deduce that

$$\mathcal{NZ}(n,\theta) \ge B(a_d)^{-1/d} n^{1/d} - c_4,$$

as asserted. □

We may also ask for a finer measure of complexity than simply counting the number of non-zero digits and consider the number of digit changes.

For an integer $b \ge 2$, an irrational real number ξ whose b-ary expansion is given by (2.1), and a positive integer n, we set

$$\mathcal{NBDC}(n,\xi,b) := \#\{\ell : 1 \le \ell \le n, a_\ell \ne a_{\ell+1}\},$$

which counts the number of digits followed by a different digit, among the first n digits in the b-ary expansion of ξ. The functions $n \mapsto \mathcal{NBDC}(n,\xi,b)$ have been introduced in [23]. Using this notion for measuring the complexity of a real number, Theorem 13.3 below, proved in [23, 31], shows that algebraic irrational numbers are ‘not too simple’.

Theorem 13.3. *Let $b \geq 2$ be an integer. For every irrational, real algebraic number θ, there exist an effectively computable constant $n_0(\theta, b)$, depending only on θ and b, and an effectively computable constant c, depending only on the degree of θ, such that*

$$\mathcal{NBDC}(n, \theta, b) \geq c\,(\log n)^{3/2}\,(\log\log n)^{-1/2} \tag{2.67}$$

for every integer $n \geq n_0(\theta, b)$.

A weaker result than (2.67), namely that

$$\lim_{n \to +\infty} \frac{\mathcal{NBDC}(n, \theta, b)}{\log n} = +\infty, \tag{2.68}$$

follows quite easily from Ridout's Theorem 6.6. The proof of Theorem 13.3 depends on a quantitative version of Ridout's Theorem. We point out that the lower bound in (2.67) does not depend on b.

Further results on the number of non-zero digits and the number of digit changes in the b-ary expansion of algebraic numbers have been obtained by Kaneko [45, 46].

Bibliography

[1] B. Adamczewski and Y. Bugeaud, On the complexity of algebraic numbers, II. Continued fractions. *Acta Math.* 195 (2005), 1–20.

[2] B. Adamczewski and Y. Bugeaud, On the independence of expansions of algebraic numbers in an integer base. *Bull. London Math. Soc.* 39 (2007), 283–289.

[3] B. Adamczewski and Y. Bugeaud, On the complexity of algebraic numbers I. Expansions in integer bases. *Ann. of Math.* 165 (2007), 547–565.

[4] B. Adamczewski and Y. Bugeaud, On the Maillet–Baker continued fractions. *J. reine angew. Math.* 606 (2007), 105–121.

[5] B. Adamczewski and Y. Bugeaud, Palindromic continued fractions. *Ann. Inst. Fourier (Grenoble)* 57 (2007), 1557–1574.

[6] B. Adamczewski and Y. Bugeaud, Mesures de transcendance et aspects quantitatifs de la méthode de Thue–Siegel–Roth–Schmidt. *Proc. London Math. Soc.* 101 (2010), 1–31.

[7] B. Adamczewski and Y. Bugeaud, Transcendence measures for continued fractions involving repetitive or symmetric patterns. *J. Europ. Math. Soc.* 12 (2010), 883–914.

[8] B. Adamczewski and Y. Bugeaud, Diophantine approximation and transcendence. In *Encyclopedia of Mathematics and its Applications* 135, Cambridge University Press, 2010, pp. 424–465.

[9] B. Adamczewski and Y. Bugeaud, Nombres réels de complexité sous-linéaire : mesures d'irrationalité et de transcendance. *J. reine angew. Math.* 658 (2011), 65–98.

[10] B. Adamczewski, Y. Bugeaud, and F. Luca, Sur la complexité des nombres algébriques. *C. R. Acad. Sci. Paris* 339 (2004), 11–14.

[11] B. Adamczewski and C. Faverjon, Chiffres non nuls dans le développement en base entière d'un nombre algébrique irrationnel. *C. R. Math. Acad. Sci. Paris* 350 (2012), 1–4.

[12] B. Adamczewski and N. Rampersad, On patterns occurring in binary algebraic numbers. *Proc. Am. Math. Soc.* 136 (2008), 3105–3109.

[13] R. Adler, M. Keane, and M. Smorodinsky, A construction of a normal number for the continued fraction transformation. *J. Number Theory* 13 (1981), 95–105.

[14] J.-P. Allouche, Nouveaux résultats de transcendance de réels à développements non aléatoire. *Gaz. Math.* 84 (2000), 19–34.

[15] J.-P. Allouche, J. L. Davison, M. Queffélec, and L. Q. Zamboni, Transcendence of Sturmian or morphic continued fractions. *J. Number Theory* 91 (2001), 39–66.

[16] J.-P. Allouche and J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge University Press, Cambridge, 2003.

[17] D. H. Bailey, J. M. Borwein, R. E. Crandall, and C. Pomerance, On the binary expansions of algebraic numbers. *J. Théor. Nombres Bordeaux* 16 (2004), 487–518.

[18] D. H. Bailey and R. E. Crandall, Random generators and normal numbers. *Experiment. Math.* 11 (2002), 527–546.

[19] V. Becher and P. A. Heiber, On extending de Bruijn sequences. *Inform. Process Lett.* 111 (2011), 930–932.

[20] Yu. F. Bilu, The many faces of the subspace theorem [after Adamczewski, Bugeaud, Corvaja, Zannier…]. Séminaire Bourbaki. Vol. 2006/2007. *Astérisque* 317 (2008), Exp. No. 967, vii, 1–38.

[21] É. Borel, Les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. Math. Palermo* 27 (1909), 247–271.

[22] Y. Bugeaud, Approximation by Algebraic Numbers. *Cambridge Tracts in Mathematics* 160, Cambridge, 2004.

[23] Y. Bugeaud, On the b-ary expansion of an algebraic number. *Rend. Sem. Mat. Univ. Padova* 118 (2007), 217–233.

[24] Y. Bugeaud, An explicit lower bound for the block complexity of an algebraic number. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.* 19 (2008), 229–235.

[25] Y. Bugeaud, On the convergents to algebraic numbers. In *Analytic Number Theory*, 133–143, Cambridge Univ. Press, Cambridge, 2009.

[26] Y. Bugeaud, Quantitative versions of the Subspace Theorem and applications. *J. Théor. Nombres Bordeaux* 23 (2011), 35–57.

[27] Y. Bugeaud, Distribution Modulo one and Diophantine Approximation. *Cambridge Tracts in Mathematics* 193, Cambridge, 2012.

[28] Y. Bugeaud, Continued fractions with low complexity: Transcendence measures and quadratic approximation. *Compos. Math.* 148 (2012), 718–750.

[29] Y. Bugeaud, Automatic continued fractions are transcendental or quadratic. *Ann. Sci. École Norm. Sup.* 46 (2013), 1005–1022.

[30] Y. Bugeaud, Transcendence of stammering continued fractions. In *Number Theory and Related Fields: In Memory of Alf van der Poorten*, J. M. Borwein et al. (eds.), Springer Proceedings in Mathematics & Statistics 43, 2013.

[31] Y. Bugeaud and J.-H. Evertse, On two notions of complexity of algebraic numbers. *Acta Arith.* 133 (2008), 221–250.

[32] J. Cassaigne, Sequences with grouped factors. In *DLT'97, Developments in Language Theory* III, Thessaloniki, Aristotle University of Thessaloniki, 1998, pp. 211–222.

[33] J. W. S. Cassels, *An Introduction to Diophantine Approximation*. Cambridge Tracts in Math. and Math. Phys., vol. 99, Cambridge University Press, 1957.

[34] D. G. Champernowne, The construction of decimals normal in the scale of ten. *J. London Math. Soc.* 8 (1933), 254–260.

[35] A. H. Copeland and P. Erdős, Note on normal numbers. *Bull. Am. Math. Soc.* 52 (1946), 857–860.

[36] K. Dajani and C. Kraaikamp, *Ergodic Theory of Numbers*. Carus Mathematical Monographs, 29. Mathematical Association of America, Washington, DC, 2002. x+190 pp.

[37] H. Davenport and P. Erdős, Note on normal decimals. *Canadian J. Math.* 4 (1952), 58–63.

[38] H. Davenport and K. F. Roth, Rational approximations to algebraic numbers. *Mathematika* 2 (1955), 160–167.

[39] L. Euler, De fractionibus continuis. *Commentarii Acad. Sci. Imperiali Petropolitanae* 9 (1737).

[40] J.-H. Evertse, The number of algebraic numbers of given degree approximating a given algebraic number. In *Analytic Number Theory* (Kyoto, 1996), 53–83, London Math. Soc. Lecture Note Ser. 247, Cambridge Univ. Press, Cambridge, 1997.

[41] J.-H. Evertse and R. G. Ferretti, A further improvement of the Quantitative Subspace Theorem. *Ann. of Math.* 177 (2013), 513–590.

[42] J.-H. Evertse and H. P. Schlickewei, A quantitative version of the Absolute Subspace Theorem. *J. reine angew. Math.* 548 (2002), 21–127.

[43] S. Ferenczi and Ch. Mauduit, Transcendence of numbers with a low complexity expansion. *J. Number Theory* 67 (1997), 146–161.

[44] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 5th. edition, Clarendon Press, 1979.

[45] H. Kaneko, On the binary digits of algebraic numbers. *J. Austral. Math. Soc.* 89 (2010), 233–244.

[46] H. Kaneko, On the number of digit changes in base-b expansions of algebraic numbers. *Uniform Distribution Theory* 7 (2012), 141–168.

[47] A. Ya. Khintchine, *Continued Fractions*. The University of Chicago Press, Chicago Ill., London, 1964.

[48] M. J. Knight, An 'Ocean of Zeroes' proof that a certain non-Liouville number is transcendental. *Am. Math. Monthly* 98 (1991), 947–949.

[49] A. N. Korobov, Continued fractions of some normal numbers. *Mat. Zametki* 47 (1990), 28–33, 158 (in Russian). English transl. in Math. Notes 47 (1990), 128–132.

[50] J. L. Lagrange, *Additions au mémoire sur la résolution des équations numériques*, Mém. Berl. 24 (1770).

[51] J. Liouville, Remarques relatives à des classes très-étendues de quantités dont la valeur n'est ni algébrique, ni même réductible à des irrationnelles algébriques, *C. R. Acad. Sci. Paris* 18 (1844), 883–885.

[52] J. Liouville, Nouvelle démonstration d'un théorème sur les irrationnelles algébriques, *C. R. Acad. Sci. Paris* 18 (1844), 910–911.

[53] K. Mahler, Some suggestions for further research. *Bull. Austral. Math. Soc.* 29 (1984), 101–108.

[54] G. Martin, Absolutely abnormal numbers. *Am. Math. Monthly* 108 (2001), 746–754.

[55] M. Morse and G. A. Hedlund, Symbolic dynamics. *Am. J. Math.* 60 (1938), 815–866.

[56] M. Morse and G. A. Hedlund, Symbolic dynamics II. *Am. J. Math.* 62 (1940), 1–42.

[57] O. Perron, *Die Lehre von den Ketterbrüchen*, Teubner, Leipzig, 1929.

[58] A. J. van der Poorten, An introduction to continued fractions. In *Diophantine Analysis* (Kensington, 1985), 99–138, London Math. Soc. Lecture Note Ser. 109, Cambridge Univ. Press, Cambridge, 1986.

[59] D. Ridout, Rational approximations to algebraic numbers. *Mathematika* 4 (1957), 125–131.

[60] T. Rivoal, On the bits counting function of real numbers. *J. Austral. Math. Soc.* 85 (2008), 95–111.

[61] K. F. Roth, Rational approximations to algebraic numbers. *Mathematika* 2 (1955), 1–20; corrigendum, 168.

[62] W. M. Schmidt, Simultaneous approximations to algebraic numbers by rationals. *Acta Math.* 125 (1970), 189–201.

[63] W. M. Schmidt, Norm form equations. *Ann. of Math.* 96 (1972), 526–551.

[64] W. M. Schmidt, Diophantine Approximation, *Lecture Notes in Math.* 785, Springer, Berlin, 1980.

[65] W. M. Schmidt, The subspace theorem in Diophantine approximation. *Compositio Math.* 69 (1989), 121–173.

[66] R. G. Stoneham, On the uniform ε-distribution of residues within the periods of rational fractions with applications to normal numbers. *Acta Arith.* 22 (1973), 371–389.

[67] A. Thue, Über Annäherungswerte algebraischer Zahlen. *J. reine angew. Math.* 135 (1909), 284–305.

[68] G. Troi and U. Zannier, Note on the density constant in the distribution of self-numbers. II. *Boll. Unione Mat. Ital. Sez. B Artic. Ric. Mat.* (8) 2 (1999), 397–399.

[69] U. Zannier, *Some Applications of Diophantine Approximation to Diophantine Equations*, Editrice Forum, Udine, 2003.

Chapter 3

Multiplicative Toeplitz matrices and the Riemann zeta function

Titus Hilberdink

Contents

1 Introduction

In this short course, we aim to highlight connections between a certain class of matrices and Dirichlet series, in particular the Riemann zeta function. The matrices we study are of the form

$$\begin{pmatrix} f(1) & f(\frac{1}{2}) & f(\frac{1}{3}) & f(\frac{1}{4}) & \cdots \\ f(2) & f(1) & f(\frac{2}{3}) & f(\frac{1}{2}) & \cdots \\ f(3) & f(\frac{3}{2}) & f(1) & f(\frac{3}{4}) & \cdots \\ f(4) & f(2) & f(\frac{4}{3}) & f(1) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \qquad (*)$$

i.e., with entries $a_{ij} = f(i/j)$ for some function $f\colon \mathbb{Q}^+ \to \mathbb{C}$. They are a multiplicative version of Toeplitz matrices which have entries of the form $a_{ij} = a_{i-j}$. For this reason we call them *Multiplicative Toeplitz Matrices*.

Toeplitz matrices (and operators) have been studied in great detail by many authors. They are most naturally studied by associating with them a function (or 'symbol') whose Fourier coefficients make up the matrix. With $a_{ij} = a_{i-j}$, this 'symbol' is

$$a(t) = \sum_{n=-\infty}^{\infty} a_n t^n. \qquad t \in \mathbb{T}$$

Then properties of the matrix (or rather the operator induced by the matrix) imply properties of the symbol and vice versa. For example, the boundedness of the operator is essentially related to the boundedness of the symbol, while invertibility of the operator is closely related to $a(t)$ not vanishing on the unit circle.

For matrices of the form ($*$) we associate, by analogy, the (formal) series

$$\sum_{q\in\mathbb{Q}^+} f(q)q^{it},$$

where q ranges over the positive rationals. Note, in particular, that if f is supported on the natural numbers, this becomes the Dirichlet series

$$\sum_{n\in\mathbb{N}} f(n)n^{it}.$$

In the special case where $f(n) = n^{-\alpha}$, the symbol becomes $\zeta(\alpha - it)$. It is quite natural then to ask to what extent properties of these Multiplicative Toeplitz Matrices are related to properties of the associated symbol. Rather surprisingly perhaps, these type of matrices do not appear to have been studied much at all – at least not in this respect. Finite truncations of them have appeared on occasions, notably Redheffer's matrix [29], the determinant of which is related to the Riemann Hypothesis. Denoting by $A_n(f)$ the $n \times n$ matrix with entries $f(i/j)$ if $j|i$ and zero otherwise, it is easy to see that

$$A_n(f)A_n(g) = A_n(f * g), \qquad \text{where } f * g \text{ is Dirichlet convolution,}$$

since the ij^{th} entry on the left product is

$$\sum_{r=1}^{n} A_n(f)_{ir}A_n(g)_{rj} = \sum_{j|r|i} f\Big(\frac{i}{r}\Big)g\Big(\frac{r}{j}\Big) = \sum_{d|i/j} f\Big(\frac{i/j}{d}\Big)g(d)$$

if $j|i$ by putting $r = jd$, and zero otherwise. With 1 and μ denoting the constant 1 and the Möbius functions, respectively, it follows that $A_n(1)A_n(\mu) = I_n$ – the identity matrix. Note also that $\det A_n(1) = \det A_n(\mu) = 1$. Redheffer's matrix is

$$R_n = A_n(1) + \begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} = A_n(1) + E_n,$$

say, where the matrix E_n has only 1s on the topmost row from the 2nd column onwards. Then, with $M(n) = \sum_{r=1}^{n} \mu(r)$,

$$R_n A_n(\mu) = I_n + \begin{pmatrix} M(n)-1 & * & \cdots & * \\ 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} = \begin{pmatrix} M(n) & * & \cdots & * \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & 1 \end{pmatrix},$$

so that $\det R_n = M(n)$. The well-known connection between the Riemann Hypothesis (RH) and $M(n)$ therefore implies that RH holds if and only if $\det R_n = O(n^{\frac{1}{2}+\varepsilon})$ for every $\varepsilon > 0$. (See also [20] for estimates of the largest eigenvalue of R_n).

Briefly then, the course is designed as follows: In Section 2, we recall some basic aspects of the theory of Toeplitz operators, in particular their boundedness and invertibility. In Section 3, we study bounded multiplicative Toeplitz operators. This is partly based on some of Toeplitz's own work [34], [35] and recent results from [17] and [18], but we also present new results, mainly in Section 3. Thus Theorem 3.1 is new, generalising Theorem 2.1 of [18], which in turn is now contained in Corollary 3.2. Also Subsection 3.2 and parts of 3.4 are new.

Preliminaries and Notation

(a) The sequence spaces l^p $(1 \le p < \infty)$ consist of sequences (a_n) for which $\sum_{n=1}^{\infty} |a_n|^p$ converges. They are Banach spaces with the norm

$$\|(a_n)\|_p = \left(\sum_{n=1}^{\infty} |a_n|^p\right)^{1/p}.$$

The space l^∞ is the space of all bounded sequences, equipped with the norm $\|(a_n)\|_\infty = \sup_{n\in\mathbb{N}} |a_n|$. We shall also use $l^p(\mathbb{Q}^+)$, which is the space of sequences a_q where q ranges over the positive rationals such that $\sum_q |a_q|^p < \infty$, with analogous norms and also for $p = \infty$.

l^2 and $l^2(\mathbb{Q}^+)$ are Hilbert spaces with the inner products

$$\langle a, b\rangle = \sum_{n=1}^{\infty} a_n \overline{b_n} \quad \text{and} \quad \langle a, b\rangle = \sum_{q\in\mathbb{Q}^+} a_q \overline{b_q},$$

respectively.

(b) Let $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$ – the unit circle. We denote by $L^2(\mathbb{T})$ the space of square-integrable functions on $\mathbb{T}$. $L^2(\mathbb{T})$ is a Hilbert space with the inner product and corresponding norm given by

$$\langle f, g\rangle = \int_{\mathbb{T}} f\overline{g} = \frac{1}{2\pi}\int_0^{2\pi} f(e^{i\theta})\overline{g(e^{i\theta})}\, d\theta, \qquad \|f\| = \sqrt{\int_{\mathbb{T}} |f|^2}.$$

The space $L^\infty(\mathbb{T})$ consists of the essentially bounded functions on $\mathbb{T}$ with norm $\|f\|_\infty$ denoting the essential supremum of f. (Strictly speaking, L^2 and L^∞ consist of *equivalence classes* of functions satisfying the appropriate conditions, with two functions belonging to the same class if they differ on a set of measure zero.)

Let $\chi_n(t) = t^n$ for $n \in \mathbb{Z}$. Then $(\chi_n)_{n\in\mathbb{Z}}$ is an orthonormal basis in $L^2(\mathbb{T})$ and $L^2(\mathbb{T})$ is isometrically isomorphic to $l^2(\mathbb{Z})$ via the mapping $f \mapsto (f_n)_{n\in\mathbb{Z}}$, where f_n are the *Fourier coefficients* of f, i.e.,

$$f_n = \langle f, \chi_n \rangle = \int_{\mathbb{T}} f\overline{\chi_n}.$$

(c) A linear operator φ on a Banach space X is *bounded* if $\|\varphi x\| \le C\|x\|$ for all $x \in X$. In this case the *operator norm* of φ is defined to be

$$\|\varphi\| = \sup_{x\in X, x\neq 0} \frac{\|\varphi x\|}{\|x\|} = \sup_{\|x\|=1} \|\varphi x\|.$$

The algebra of bounded linear operators on X is denoted by $B(X)$.

(d) An infinite matrix $A = (a_{ij})$ induces a bounded operator on a Hilbert space H if there exists $\varphi \in B(H)$ such that

$$a_{ij} = \langle \varphi e_j, e_i \rangle,$$

where (e_i) is an orthonormal basis of H. Note that not every infinite matrix induces a bounded operator, and it may be difficult to tell when it does.

(e) For the later sections we require the usual $O, o, \sim, \ll$ notation. Given f, g defined on neighbourhoods of ∞ with g eventually positive, we write $f(x) = o(g(x))$ (or simply $f = o(g)$) to mean $\lim_{x\to\infty} f(x)/g(x) = 0$, $f(x) = O(g(x))$ to mean $|f(x)| \le Ag(x)$ for some constant A and all x sufficiently large, and $f(x) \sim g(x)$ to mean $\lim_{x\to\infty} f(x)/g(x) = 1$.

The notation $f \ll g$ means the same as $f = O(g)$, while $f \lesssim g$ means $f(x) \le (1 + o(1))g(x)$.

2 Toeplitz matrices and operators – a brief overview

Toeplitz matrices are matrices of the form

$$T = \begin{pmatrix} a_0 & a_{-1} & a_{-2} & a_{-3} & \cdots \\ a_1 & a_0 & a_{-1} & a_{-2} & \cdots \\ a_2 & a_1 & a_0 & a_{-1} & \cdots \\ a_3 & a_2 & a_1 & a_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{3.1}$$

i.e., $T = (t_{ij})$, where $t_{ij} = a_{i-j}$. They are characterised by being constant on diagonals.

For a Toeplitz matrix, the question of boundedness of T was solved by Toeplitz.

Theorem 2.1 (Toeplitz [34]). *The matrix T induces a bounded operator on l^2 if and only if there exists $a \in L^\infty(\mathbb{T})$ whose Fourier coefficients are a_n $(n \in \mathbb{Z})$. If this is the case, then $\|T\| = \|a\|_\infty$.*

We refer to the function a as the 'symbol' of the matrix T, and we write $T(a)$.

Sketch of Proof. For $a \in L^2(\mathbb{T})$, the multiplication operator

$$M(a)\colon L^2(\mathbb{T}) \longrightarrow L^2(\mathbb{T}),\ f \longmapsto af$$

is bounded if and only if $a \in L^\infty(\mathbb{T})$. If bounded, then $\|M(a)\| = \|a\|_\infty$. The matrix representation of $M(a)$ with respect to $(\chi_n)_{n\in\mathbb{Z}}$ is given by

$$\langle M(a)\chi_j, \chi_i\rangle = \langle a\chi_j, \chi_i\rangle = \int_{\mathbb{T}} a\chi_j\overline{\chi_i} = \int_{\mathbb{T}} a\overline{\chi_{i-j}} = a_{i-j},$$

i.e., by the so-called *Laurent matrix*

$$L(a) := \left(\begin{array}{ccc|cccc} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & a_0 & a_{-1} & a_{-2} & a_{-3} & a_{-4} & \cdots \\ \cdots & a_1 & a_0 & a_{-1} & a_{-2} & a_{-3} & \cdots \\ \hline \cdots & a_2 & a_1 & a_0 & a_{-1} & a_{-2} & \cdots \\ \cdots & a_3 & a_2 & a_1 & a_0 & a_{-1} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{array}\right). \tag{3.2}$$

The matrix for T is just the lower right quarter of $L(a)$. We can therefore think of T as the compression $PL(a)P$, where P is the projection of $l^2(\mathbb{Z})$ onto $l^2 = l^2(\mathbb{N})$. An easy argument shows that T is bounded if and only if $a \in L^\infty$, and then $\|T\| = \|L(a)\| = \|a\|_\infty$. □

Hardy space. Let $H^2(\mathbb{T})$ denote the subspace of $L^2(\mathbb{T})$ of functions f whose Fourier coefficients f_n vanish for $n < 0$. Let P be the orthogonal projection of L^2 onto H^2, i.e., $P(\sum_{n\in\mathbb{Z}} f_n\chi_n) = \sum_{n\geq 0} f_n\chi_n$. The operator $f \mapsto P(af)$ has matrix representation (3.1). For, with $j \geq 0$ (so that $\chi_j \in H^2(\mathbb{T})$),

$$\langle T(a)\chi_j, \chi_i\rangle = \int_{\mathbb{T}} P(a\chi_j)\overline{\chi_i} = \int_{\mathbb{T}} P\Big(\sum_{n\in\mathbb{Z}} a_n\chi_{n+j}\Big)\chi_{-i} = \sum_{n\geq 0} a_{n-j}\int_{\mathbb{T}}\chi_{n-i} = a_{i-j}$$

if $i \geq 0$, and zero otherwise. Hence, we can equivalently view $T(a)$ as the operator

$$T(a)\colon H^2(\mathbb{T}) \longrightarrow H^2(\mathbb{T}),\ f \longmapsto P(af).$$

2.1 $C(\mathbb{T})$, $W(\mathbb{T})$, and winding number Let $C(\mathbb{T})$ denote the space of continuous functions on $\mathbb{T}$. For $a \in C(\mathbb{T})$ such that $a(t) \neq 0$ for all $t \in \mathbb{T}$, we denote by $\text{wind}(a,0)$ the *winding number* of a with respect to zero. More generally, $\text{wind}(a,\lambda)$ $=\text{wind}(a-\lambda,0)$ denotes thewinding number with respect $\lambda \in \mathbb{C}$. For example, $\text{wind}(\chi_n,0) = n$.

The *Wiener Algebra* is the set of absolutely convergent Fourier series:

$$W(\mathbb{T}) = \left\{ \sum_{-\infty}^{\infty} a_n \chi_n : \sum_{-\infty}^{\infty} |a_n| < \infty \right\}.$$

Some properties:

(i) $W(\mathbb{T})$ forms a Banach algebra under pointwise multiplication, with norm

$$\|a\|_W := \sum_{-\infty}^{\infty} |a_n|.$$

(ii) (Wiener's Theorem) If $a \in W$ and $a(t) \neq 0$ for all $t \in \mathbb{T}$, then $a^{-1} \in W$.

(iii) If $a \in W(\mathbb{T})$ has no zeros and $\text{wind}(a,0) = 0$, then $a = e^b$ for some $b \in W(\mathbb{T})$.

We have

$$W(\mathbb{T}) \subset C(\mathbb{T}) \subset L^{\infty}(\mathbb{T}) \subset L^2(\mathbb{T}).$$

2.2 Invertibility and Fredholmness Let A be a bounded operator on a Banach space X.

(i) A is *invertible* if there exists a bounded operator B on X such that $AB = BA = I$. As such, B is the unique *inverse* of A, and we write $B = A^{-1}$. The *spectrum* of A is the set

$$\sigma(A) = \{\lambda \in \mathbb{C} : \lambda I - A \text{ is not invertible in } X\}.$$

The *kernel* and *image* of A are defined by

$$\text{Ker}\, A = \{x \in X : Ax = 0\}, \qquad \text{Im}\, A = \{Ax : x \in X\}.$$

(ii) The operator A is *Fredholm* if ImA is a closed subspace of X and both Ker$A$ and $X/\text{Im}A$ are finite-dimensional. As such, the *index* of A is defined to be

$$\text{Ind}\, A = \dim \text{Ker}\, A - \dim (X/\text{Im}\, A).$$

For example, $T(\chi_n)$ is Fredholm with Ind $T(\chi_n) = -n$.

Equivalently, A is Fredholm if it is invertible modulo compact operators; that is, there exists bounded operator B on X such that $AB - I$ and $BA - I$ are both compact.

The *essential spectrum* of A is the set

$$\sigma_{\text{ess}}(A) = \{\lambda \in \mathbb{C} : \lambda I - A \text{ is not Fredholm in } X\}.$$

Clearly $\sigma_{\text{ess}}(A) \subset \sigma(A)$. Note that A invertible implies A is Fredholm of index zero. For Toeplitz operators, the converse actually holds (see [5], p. 12).

2.3 Hankel matrices These are matrices of the form

$$H = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & \cdots \\ a_2 & a_3 & a_4 & a_5 & \cdots \\ a_3 & a_4 & a_5 & a_6 & \cdots \\ a_4 & a_5 & a_6 & a_7 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{3.3}$$

i.e., $H = (h_{ij})$, where $h_{ij} = a_{i+j-1}$. They are characterised by being constant on cross diagonals. The boundedness of H was solved by Nehari, and the compactness of H by Hartman.

Theorem 2.2 ([28], [14]). *The matrix H generates a bounded operator on l^2 if and only if there exists $b \in L^\infty(\mathbb{T})$ (with Fourier coefficients b_n) such that $b_n = a_n$ for $n \geq 1$. Furthermore, the operator H is compact if and only if $b \in C(\mathbb{T})$.*

We refer to the function a as the 'symbol' of the matrix H, and we write $H(a)$. Given a function a defined on $\mathbb{T}$, let $\tilde{a}$ be the function

$$\tilde{a}(t) = a(1/t) \quad (t \in \mathbb{T}).$$

Proposition 2.3. *For $a, b \in L^\infty(\mathbb{T})$,*

$$T(ab) = T(a)T(b) + H(a)H(\tilde{b})$$
$$H(ab) = H(a)T(\tilde{b}) + T(a)H(b).$$

Proof. The matrix $L(a)$ in (3.2) is of the form

$$L(a) = \left(\begin{array}{c|c} T(a) & H(\tilde{a}) \\ \hline H(a) & T(a) \end{array}\right).$$

Since $L(ab) = L(a)L(b)$, the result follows by multiplying the 2 × 2 matrices. □

As a special case, we see that $T(abc) = T(a)T(b)T(c)$ for $a \in \overline{H^\infty}, b \in L^\infty, c \in H^\infty$. The space H^∞ is defined analogously to L^∞. ($T(a)$ is upper-triangular and $T(c)$ is lower-triangular.)

By Theorem 2.2, if $a, b \in C(\mathbb{T})$, then $H(a)H(\tilde{b})$ is compact, so that $T(ab) - T(a)T(b)$ is compact. In particular, if a has no zeros on $\mathbb{T}$, we can take $b = a^{-1} \in C(\mathbb{T})$. Then $T(ab) = T(1) = I$, so $T(a)$ is invertible modulo compact operators (i.e., Fredholm) with 'inverse' $T(a^{-1})$. This type of reasoning leads to:

Theorem 2.4 (Gohberg [9]). *Let $a \in C(\mathbb{T})$. Then $T(a)$ is Fredholm if and only if a has no zeros on $\mathbb{T}$, in which case*

$$\operatorname{Ind} T(a) = -\operatorname{wind}(a, 0).$$

Hence $T(a)$ is invertible if and only if a has no zeros on $\mathbb{T}$ and $\mathrm{wind}(a,0) = 0$. *Equivalently, since $T(\lambda - a) = \lambda I - T(a)$ for $\lambda \in \mathbb{C}$, we have*

$$\begin{aligned}\sigma_{\mathrm{ess}}(T(a)) &= a(\mathbb{T}),\\ \sigma(T(a)) &= a(\mathbb{T}) \cup \{\lambda \in \mathbb{C} \setminus a(\mathbb{T}) : \mathrm{wind}(a,\lambda) \neq 0\}.\end{aligned}$$

Sketch of Proof. We have seen that $a \neq 0$ on $\mathbb{T}$ implies $T(a)$ is Fredholm. In this case, let $\mathrm{wind}(a,0) = k$. Then a is homotopic to χ_k, and (since the index varies continuously)

$$\mathrm{Ind}\, T(a) = \mathrm{Ind}\, T(\chi_k) = -k = -\mathrm{wind}\,(a,0).$$

For the converse, suppose $T(a)$ is Fredholm with index k, but a has zeros on $\mathbb{T}$. Then a can be slightly perturbed to produce two functions $b, c \in W(\mathbb{T})$ without zeros such that $\|a-b\|_\infty$ and $\|a-c\|_\infty$ are as small as we please, but $\mathrm{wind}(b,0)$ and $\mathrm{wind}(c,0)$ differ by one. As the index is stable under small perturbations, $T(b)$ and $T(c)$ are Fredholm with equal index. But $\mathrm{Ind}\, T(b) = -\mathrm{wind}(b,0)$ and $\mathrm{Ind}\, T(c) = -\mathrm{wind}(c,0)$ (by above), so $\mathrm{wind}(b,0) - \mathrm{wind}(c,0) = 0$ — a contradiction. □

2.4 Wiener–Hopf factorization

Since $W(\mathbb{T}) \subset C(\mathbb{T})$, Theorem 2.4 applies to $W(\mathbb{T})$. However, for Wiener symbols we can obtain a quite explicit form for the inverse when it exists. This is because Wiener functions can be factorized.

Denote by W_+ and W_- the subspaces of W consisting of functions

$$\sum_{n=0}^{\infty} a_n t^n \quad \text{and} \quad \sum_{n=0}^{\infty} a_n t^{-n} \qquad t \in \mathbb{T}$$

respectively, where $\sum |a_n| < \infty$.

Theorem 2.5 (Wiener–Hopf factorization). *Let $a \in W(\mathbb{T})$ such that a has no zeros, and let* $\mathrm{wind}(a,0) = k$. *Then there exist $a_- \in W_-$ and $a_+ \in W_+$ such that*

$$a = \chi_k a_- a_+.$$

Proof. We have $\mathrm{wind}(a\chi_{-k},0) = 0$. So $a\chi_{-k} = e^b$ for some $b \in W$. But $b = b_- + b_+$, where $b_- \in W_-$ and $b_+ \in W_+$. Hence, writing $a_- = e^{b-}$ and $a_+ = e^{b+}$ gives

$$a\chi_{-k} = e^{b-}e^{b+} = a_-a_+. \qquad \square$$

Theorem 2.6 (Krein [23]). *Let $a \in W(\mathbb{T})$. Then $T(a)$ is Fredholm if and only if a has no zeros on $\mathbb{T}$, in which case*

$$\mathrm{Ind}\, T(a) = -\mathrm{wind}(a,0).$$

In particular, $T(a)$ is invertible if and only if a has no zeros on $\mathbb{T}$ and $\mathrm{wind}(a,0) = 0$. *In this case*

$$T(a)^{-1} = T(a_+^{-1})T(a_-^{-1}),$$

where $a = a_+a_-$ is the Wiener–Hopf factorization of a.

Proof of second part. Note that if $a \in W_-$, then $H(a) = 0$, while if $a \in W_+$, then $H(\tilde{a}) = 0$. Suppose a has no zeros on $\mathbb{T}$ and $\operatorname{wind}(a, 0) = 0$. Then a factorizes as $a = a_- a_+$ with $a_\pm \in W_\pm$. Applying Proposition 2.3 with a_- and a_+ in turn gives

$$T(a_-^{-1})T(a_-) = T(a_-^{-1}a_-) = I = T(a_-a_-^{-1}) = T(a_-)T(a_-^{-1}),$$
$$T(a_+^{-1})T(a_+) = T(a_+^{-1}a_+) = I = T(a_+a_+^{-1}) = T(a_+)T(a_+^{-1}),$$

so that $T(a_\pm)$ are invertible with $T(a_\pm)^{-1} = T(a_\pm^{-1})$. But also $T(a) = T(a_-a_+) = T(a_-)T(a_+)$ (by Proposition 2.3). Hence $T(a)^{-1} = T(a_+)^{-1}T(a_-)^{-1} = T(a_+^{-1})T(a_-^{-1})$. □

3 Bounded multiplicative Toeplitz matrices and operators

3.1 Criterion for boundedness on l^2 Now we consider the linear operators induced by matrices of the form $(*)$, regarding them as operators on sequence spaces, in particular l^2.

For a function $f : \mathbb{Q}^+ \to \mathbb{C}$ on the positive rationals, we define

$$\sum_{q \in \mathbb{Q}^+} f(q) = \lim_{N \to \infty} \sum_{\substack{m,n \le N \\ (m,n)=1}} f\left(\frac{m}{n}\right), \quad \text{whenever this limit exists.}$$

We shall sometimes abbreviate the left-hand sum by $\sum_q f(q)$. We say that $f \in l^1(\mathbb{Q}^+)$ if

$$\sum_{q \in \mathbb{Q}^+} |f(q)|$$

converges. In this case, the function

$$F(t) = \sum_{q \in \mathbb{Q}^+} f(q) q^{it} \qquad t \in \mathbb{R}$$

exists and is uniformly continuous on $\mathbb{R}$. Note that for $\lambda > 0$,

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} F(t) \lambda^{-it}\, dt = \begin{cases} f(q), & \text{if } \lambda = q \in \mathbb{Q}^+, \\ 0, & \text{otherwise.} \end{cases} \tag{3.4}$$

Theorem 3.1. *Let $f \in l^1(\mathbb{Q}^+)$ and let φ_f denote the mapping $(a_n) \mapsto (b_n)$ where*

$$b_n = \sum_{m=1}^{\infty} f\left(\frac{n}{m}\right) a_m.$$

Then φ_f is bounded on l^2 with operator norm

$$\|\varphi_f\| = \sup_{t\in\mathbb{R}}\bigg|\sum_{q\in\mathbb{Q}^+} f(q)q^{it}\bigg| = \|F\|_\infty.$$

Proof. We shall first prove that φ_f is bounded on l^2, showing $\|\varphi_f\| \leq \|F\|_\infty$ in the process, and then show that $\|F\|_\infty$ is also a lower bound.

For $q \in \mathbb{Q}^+$ and $N \in \mathbb{N}$, let

$$b_q^{(N)} = \sum_{m=1}^{N} f\Big(\frac{q}{m}\Big)a_m \quad \text{and} \quad b_q = \sum_{m=1}^{\infty} f\Big(\frac{q}{m}\Big)a_m.$$

Note that $b_q^{(N)} \to b_q$ as $N \to \infty$ for every $q \in \mathbb{Q}^+$, whenever a_n is bounded. We have the following formulae:

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} F(t)\bigg|\sum_{n=1}^{N} a_n n^{it}\bigg|^2 dt = \sum_{n=1}^{N}\overline{a_n}b_n^{(N)} \tag{3.5}$$

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\bigg|F(t)\sum_{n=1}^{N} a_n n^{it}\bigg|^2 dt = \sum_{q\in\mathbb{Q}^+}|b_q^{(N)}|^2. \tag{3.6}$$

(These hold for a_n bounded.) To prove these expand the integrand in a Dirichlet series. For the first formula we have

$$\frac{1}{2T}\int_{-T}^{T} F(t)\bigg|\sum_{n=1}^{N} a_n n^{it}\bigg|^2 dt = \sum_{m,n\leq N} a_m\overline{a_n}\frac{1}{2T}\int_{-T}^{T} F(t)\Big(\frac{n}{m}\Big)^{-it} dt$$

$$\longrightarrow \sum_{m,n\leq N} a_m\overline{a_n} f\Big(\frac{n}{m}\Big) = \sum_{n=1}^{N}\overline{a_n}b_n^{(N)}$$

as $T \to \infty$. For the second formula, note first that

$$F(t)\sum_{n=1}^{N} a_n n^{it} = \sum_{q\in\mathbb{Q}^+, n\leq N} f(q)a_n(qn)^{it}$$

$$= \sum_{r\in\mathbb{Q}^+}\bigg(\sum_{n\leq N} f\Big(\frac{r}{n}\Big)a_n\bigg)r^{it} = \sum_{r\in\mathbb{Q}^+} b_r^{(N)} r^{it},$$

the series converging absolutely. Thus

$$\frac{1}{2T}\int_{-T}^{T}\bigg|F(t)\sum_{n=1}^{N} a_n n^{it}\bigg|^2 dt = \sum_{q_1,q_2\in\mathbb{Q}^+} b_{q_1}^{(N)}\overline{b_{q_2}^{(N)}}\frac{1}{2T}\int_{-T}^{T}\Big(\frac{q_1}{q_2}\Big)^{it} dt \longrightarrow \sum_{q\in\mathbb{Q}^+}|b_q^{(N)}|^2$$

as $T \to \infty$.

Since $|F(t)| \leq \|F\|_\infty$, we have

$$\sum_{q\in\mathbb{Q}^+} |b_q^{(N)}|^2 \leq \|F\|_\infty^2 \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T}\left|\sum_{n=1}^{N} a_n n^{it}\right|^2 dt = \|F\|_\infty^2 \sum_{n=1}^{N} |a_n|^2.$$

Thus if $a = (a_n) \in l^2$, we have $\sum_{n=1}^{\infty} |b_n^{(N)}|^2 \leq \|F\|_\infty^2 \|a\|^2$ for every N. Letting $N \to \infty$ shows that $(b_n) \in l^2$ too (indeed $(b_q) \in l^2(\mathbb{Q}^+)$), and so φ_f is bounded on l^2, with

$$\|\varphi_f\| \leq \|F\|_\infty.$$

Now we need a lower bound. By Cauchy–Schwarz,

$$\left|\sum_{n=1}^{\infty} \overline{a_n} b_n\right|^2 \leq \sum_{n=1}^{\infty} |a_n|^2 \cdot \sum_{n=1}^{\infty} |b_n|^2.$$

Thus $\|\varphi_f\| \geq |\sum_{n=1}^{\infty} \overline{a_n} b_n|$ for every $a = (a_n) \in l^2$ with $\|a\| = 1$. Choose a_n as follows: let $N \in \mathbb{N}$ (to be determined later) and put

$$a_n = \frac{n^{-it}}{\sqrt{d(N)}} \qquad \text{for } n|N, \text{ and zero otherwise.}$$

Here $d(N)$ is the number of divisors of N. Thus $(a_n) \in l^2$ and $\|a\| = 1$. With this choice,

$$b_n = \frac{1}{\sqrt{d(N)}} \sum_{m|N} f\Big(\frac{n}{m}\Big) m^{-it} \qquad (= b_n^{(N)})$$

and so

$$\sum_{n=1}^{\infty} \overline{a_n} b_n = \frac{1}{d(N)} \sum_{n|N} n^{it} \sum_{m|N} f\Big(\frac{n}{m}\Big) m^{-it} = \frac{1}{d(N)} \sum_{m,n|N} f\Big(\frac{n}{m}\Big)\Big(\frac{n}{m}\Big)^{it}$$
$$= \frac{1}{d(N)} \sum_{q\in\mathbb{Q}^+} f(q) q^{it} S_q(N),$$

where

$$S_q(N) = \sum_{\substack{m,n|N \\ \frac{n}{m}=q}} 1.$$

Put $q = \frac{k}{l}$, where $(k,l) = 1$. Then $\frac{n}{m} = \frac{k}{l}$ if and only if $ln = km$. Since $(k,l) = 1$, this forces $k|n$ and $l|m$. So, for a contribution to the sum, we need $k, l|N$, i.e., $kl|N$. Suppose therefore that $kl|N$. Then

$$S_q(N) = \sum_{\substack{m,n|N \\ ln=km}} 1 = \sum_{rk,rl|N} 1 \qquad m = rl, n = rk \text{ with } r \in \mathbb{N}$$
$$= \sum_{r|\frac{N}{kl}} 1 = d\Big(\frac{N}{kl}\Big).$$

Writing $|q| = kl$ whenever $q = \frac{k}{l}$ in its lowest terms, gives

$$\sum_{n=1}^{\infty} \overline{a_n} b_n = \sum_{\substack{q \in \mathbb{Q}^+ \\ |q| | N}} f(q) q^{it} \frac{d(N/|q|)}{d(N)}. \tag{3.7}$$

The idea is now to choose N in such a way that it has all 'small' divisors while $\frac{d(N/|q|)}{d(N)}$ is close to 1 for all such small divisors $|q|$. Take N of the form

$$N = \prod_{p \le P} p^{\alpha_p}, \qquad \text{where } \alpha_p = \left[\tfrac{\log P}{\log p}\right].$$

Thus every natural number up to P is a divisor of N. Every q such that $|q||N$ is of the form $|q| = \prod_{p \le P} p^{\beta_p}$ $(0 \le \beta_p \le \alpha_p)$, so that

$$\frac{d(N/|q|)}{d(N)} = \prod_{p \le P} \Big(1 - \frac{\beta_p}{\alpha_p + 1}\Big).$$

If we take $|q| \le \sqrt{\log P}$, then $p^{\beta_p} \le \sqrt{\log P}$ for every prime divisor p of $|q|$. Hence, for such p, $\beta_p \le \frac{\log\log P}{2\log p}$ and $\beta_p = 0$ if $p > \sqrt{\log P}$. Thus

$$\frac{d(N/|q|)}{d(N)} = \prod_{p \le \sqrt{\log P}} \Big(1 - \frac{\beta_p}{\alpha_p + 1}\Big) \ge \prod_{p \le \sqrt{\log P}} \Big(1 - \frac{\log\log P}{2\log P}\Big)$$
$$= \Big(1 - \frac{\log\log P}{2\log P}\Big)^{\pi(\sqrt{\log P})},$$

where $\pi(x)$ is the number of primes up to x. Since $\pi(x) = O(\frac{x}{\log x})$, it follows that for all P sufficiently large, the RHS above is at least

$$1 - \frac{A}{\sqrt{\log P}}$$

for some constant A.

Let $\varepsilon > 0$. Then there exists n_0 such that $\sum_{|q| > n_0} |f(q)| < \varepsilon$. Choose $P \ge e^{n_0^2}$ so that $\sqrt{\log P} \ge n_0$. Then the modulus of the sum in (3.7) can be made as close to $\|F\|_\infty$ as we please by increasing P, for it is at least

$$\Big|\sum_{|q| \le \sqrt{\log P}} f(q) q^{it}\Big| - \frac{A}{\sqrt{\log P}} \sum_{|q| \le \sqrt{\log P}} |f(q)| - \sum_{|q| > \sqrt{\log P}} |f(q)|$$
$$\ge \Big|\sum_{q \in \mathbb{Q}^+} f(q) q^{it}\Big| - \frac{A}{\sqrt{\log P}} \sum_{q \in \mathbb{Q}^+} |f(q)| - 2 \sum_{|q| > \sqrt{\log P}} |f(q)|$$
$$> |F(t)| - \frac{A'}{\sqrt{\log P}} - 2\varepsilon,$$

where ε can be made as small as we please by making P large. Since this holds for any t, we can choose t to make $F(t)$ as close as we like to $\|F\|_\infty$. Hence $\|\varphi_f\| \geq \|F\|_\infty$ and so we must have equality. □

In the special case where $f \geq 0$, the supremum of $|F(t)|$ is attained when $t = 0$, in which case $\|F\|_\infty = F(0) = \|f\|_{1,\mathbb{Q}^+}$. Thus:

Corollary 3.2. *Let $f : \mathbb{Q}^+ \to \mathbb{C}$ such that $f \geq 0$. Then φ_f is bounded on l^2 if and only if $f \in l^1(\mathbb{Q}^+)$, in which case $\|\varphi_f\| = \|f\|_{1,\mathbb{Q}^+}$.*

Example. Take $f(n) = n^{-\alpha}$ for $n \in \mathbb{N}$, and zero otherwise. Then $F(t) = \zeta(\alpha - it)$. We shall denote φ_f by φ_α in this case. Applying Corollary 3.2, we see that φ_α is bounded on l^2 if and only if $\alpha > 1$, and the norm is $\zeta(\alpha)$.

3.2 Viewing φ_f as an operator on function spaces; Besicovitch space

We can view φ_f as an operator on functions rather than sequences. For this we need to construct the appropriate spaces.

Let A denote the set of trigonometric polynomials; i.e., the elements of A are all finite sums of the form

$$\sum_{k=1}^{n} a_k e^{i\lambda_k t},$$

where $a_k \in \mathbb{C}$ and $\lambda_k \in \mathbb{R}$. The space A has an inner product and a norm given by

$$\langle f, g\rangle = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f\overline{g} \qquad \text{and} \qquad \|f\| = \sqrt{\langle f, f\rangle} = \sqrt{\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} |f|^2}.$$

Now let B^2 (Besicovitch space) denote the closure of A with respect to this inner product; i.e., $f \in B^2$ if $\|f - f_n\| \to 0$ as $n \to \infty$ for some $f_n \in A$. We turn B^2 into a Hilbert space by identifying f and g whenever $\|f - g\| = 0$. (See [3], Chapter II.)

Now write $\chi_\lambda(t) = \lambda^{it}$ ($\lambda > 0, t \in \mathbb{R}$) and let $\hat{f}(\lambda)$ denote the Fourier coefficient

$$\hat{f}(\lambda) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f\overline{\chi_\lambda} = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f(t)\lambda^{-it}\,dt \qquad \text{where it exists.}$$

Denote by $\mathcal{F}$ the space of locally integrable $f : \mathbb{R} \to \mathbb{C}$ such that $\hat{f}(\lambda)$ exists for all $\lambda > 0$.

(a) **Fourier coefficients and series** *For $f \in B^2$, the Fourier coefficients $\hat{f}(\lambda)$ exist and $\hat{f}(\lambda)$ is non-zero on at most a countable set, say $\{\lambda_n\}_{n\in\mathbb{N}}$. The function f has the (formal) Fourier series $\sum_{n\geq 1} \hat{f}(\lambda_n)\lambda_n^{it}$.*

(b) **Uniqueness** *$f, g \in B^2$ have the same Fourier series if and only if $\|f - g\| = 0$.*

(c) **Parseval** *For $f \in B^2$,*

$$\|f\| = \lim_{T\to\infty} \sqrt{\frac{1}{2T}\int_{-T}^{T} |f|^2} = \sqrt{\sum_{\lambda} |\hat{f}(\lambda)|^2}, \tag{3.8}$$

and, more generally, for $f, g \in B^2$,

$$\langle f, g\rangle = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f\overline{g} = \sum_{\lambda} \hat{f}(\lambda)\overline{\hat{g}(\lambda)}.$$

(d) **Riesz–Fischer Theorem** *Given $\lambda_n > 0$ and $a_n \in l^2$, there exists $f \in B^2$ such that $f(t) \sim \sum_{n\geq 1} a_n \lambda_n^{it}$.*

(e) **Criterion for membership in** B^2: *With $\mathcal{F}$ as before, if $f \in \mathcal{F}$ and Parseval's identity* (3.8) *holds, then $f \in B^2$.*

Indeed, the set of λ for which $\hat{f}(\lambda) \neq 0$ is necessarily countable and we may write this as $\{\lambda_1, \lambda_2, \ldots\}$ with $\sum_{k=1}^{\infty} |\hat{f}(\lambda_k)|^2 = \|f\|^2$. Let $f_n(t) = \sum_{k\leq n} \hat{f}(\lambda_k)\lambda_k^{it}$. Then

$$\|f - f_n\|^2 = \|f\|^2 - \|f_n\|^2 = \sum_{k>n} |\hat{f}(\lambda)|^2 \to 0 \quad \text{as } n \to \infty.$$

The analogues of the Hardy and Wiener spaces: $B^2_{\mathbb{Q}^+}$, $B^2_{\mathbb{N}}$, $W_{\mathbb{Q}^+}$, $W_{\mathbb{N}}$.

(a) Let $B^2_{\mathbb{Q}^+}$ denote the subspace of B^2 of functions with exponents $\lambda = \log q$ for some $q \in \mathbb{Q}^+$. Alternatively, start with the subset of A consisting of finite trigonometric polynomials of the form $\sum a_q \chi_q$, where q ranges over a finite subset of $\mathbb{Q}^+$, and take its closure.

(b) Let $B^2_{\mathbb{N}}$ denote the subspace of B^2 of functions with exponents $\lambda = \log n$ for some $n \in \mathbb{N}$. This is the analogue of the Hardy space.

(c) Let $W_{\mathbb{Q}^+}$ denote the set of all absolutely convergent Fourier series in $B^2_{\mathbb{Q}^+}$; i.e.

$$W_{\mathbb{Q}^+} = \Big\{ \sum_{q\in\mathbb{Q}^+} c(q)\chi_q : \sum_{q\in\mathbb{Q}^+} |c(q)| < \infty \Big\}.$$

This is the analogue of the Wiener algebra. As we saw earlier, if

$$f = \sum_{q\in\mathbb{Q}^+} c(q)\chi_q \in W_{\mathbb{Q}^+},$$

then $\hat{f}(q) = c(q)$. With pointwise addition and multiplication, $W_{\mathbb{Q}^+}$ becomes an algebra. Further, $W_{\mathbb{Q}^+}$ becomes a Banach algebra with respect to the norm

$$\|f\|_W = \sum_{q\in\mathbb{Q}^+} |\hat{f}(q)|.$$

Analogously, let $W_{\mathbb{N}}$ denote the set of absolutely convergent series $\sum_{n=1}^{\infty} a_n n^{it}$.

$B^2_{\mathbb{Q}^+}$ has $(\chi_q)_{q\in\mathbb{Q}^+}$ as an orthonormal basis while $B^2_{\mathbb{N}}$ has $(\chi_n)_{n\in\mathbb{N}}$ as an orthonormal basis. The spaces $B^2_{\mathbb{Q}^+}$ and $B^2_{\mathbb{N}}$ are isometrically isomorphic to $l^2(\mathbb{Q}^+)$ and l^2, respectively, via the mappings

$$f \overset{\tau}{\longmapsto} \{\hat{f}(q)\}_{q\in\mathbb{Q}^+} \text{ and } f \overset{\tau}{\longmapsto} \{\hat{f}(n)\}_{n\geq 1}.$$

The operator $M(f)$. Let P be the projection from $B^2_{\mathbb{Q}^+}$ to $B^2_{\mathbb{N}}$; that is,

$$P\Big(\sum_{q\in\mathbb{Q}^+} c(q)q^{it}\Big) = \sum_{n\in\mathbb{N}} c(n)n^{it}.$$

For $f \in W_{\mathbb{Q}^+}$, we define the operator $M(f)$: $B^2_{\mathbb{N}} \to B^2_{\mathbb{N}}$ by $g \mapsto P(fg)$.

The matrix representation of $M(f)$ w.r.t. $\{\chi_n\}_{n\in\mathbb{N}}$ is the multiplicative Toeplitz matrix $(\hat{f}(i/j))_{i,j\geq 1}$. Indeed, if $f = \sum_q \hat{f}(q)\chi_q$, then

$$\begin{aligned} P(f\chi_j) &= P\Big(\sum_q \hat{f}(q)\chi_q\chi_j\Big) = P\Big(\sum_q \hat{f}(q)\chi_{qj}\Big) \\ &= P\Big(\sum_q \hat{f}(q/j)\chi_q\Big) = \sum_{n=1}^{\infty} \hat{f}(n/j)\chi_n. \end{aligned}$$

Hence

$$\langle M(f)\chi_j, \chi_i\rangle = \langle P(f\chi_j), \chi_i\rangle = \sum_{n=1}^{\infty} \hat{f}(n/j)\langle\chi_n, \chi_i\rangle = \hat{f}\Big(\frac{i}{j}\Big).$$

In terms of the operator φ,

$$M(f) = \tau^{-1}\varphi_{\hat{f}}\tau.$$

3.3 Interlude on $\zeta(s)$

Since this work concerns connections to Dirichlet series and the Riemann zeta function in particular, we recall a few relevant facts regarding $\zeta(s)$.

The *Riemann zeta function* is defined for $\Re s > 1$ by the Dirichlet series

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

In this half-plane $\zeta(s)$ is holomorphic and there is an analytic continuation to the whole of $\mathbb{C}$ except for a simple pole at $s = 1$ with residue 1. Furthermore, $\zeta(s)$ satisfies the *functional equation*

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

where

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

The connection to prime numbers comes from Euler's remarkable product formula

$$\zeta(s) = \prod_p \frac{1}{1-p^{-s}} \qquad \Re s > 1$$

The order of $\zeta(s)$. Considering $\zeta(\sigma + it)$ as a function of the real variable t for fixed (but arbitrary) σ, it is known that for $|t|$ large

$$\zeta_\sigma(t) \overset{\text{def}}{=} \zeta(\sigma + it) = O(|t|^A) \qquad \text{for some } A.$$

The infimum of such A is the *order* of ζ and is called the *Lindelöf* function; i.e.,

$$\mu(\sigma) = \inf\{A : \zeta(\sigma + it) = O(|t|^A)\}.$$

From the general theory of functions it is known that the Lindelöf function is convex and decreasing. Since ζ_σ is bounded for $\sigma > 1$, but $\zeta_\sigma \not\to 0$, it follows that $\mu(\sigma) = 0$ for $\sigma \geq 1$. By the functional equation and continuity of μ we then have

$$\mu(\sigma) = \begin{cases} 0, & \text{if } \sigma \geq 1, \\ \frac{1}{2} - \sigma, & \text{if } \sigma \leq 0. \end{cases} \tag{$\diamond$}$$

For $0 < \sigma < 1$, the value of $\mu(\sigma)$ is not known, but it is conjectured that the two line segments in ($\diamond$) above extend to $\sigma = \frac{1}{2}$. This is the *Lindelöf Hypothesis*. It is equivalent to the statement that

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

The Lindelöf Hypothesis is a major open problem and is a consequence of the Riemann Hypothesis, which states that $\zeta(s) \neq 0$ for $\sigma > \frac{1}{2}$.

Upper and lower bounds for ζ_σ. Let

$$Z_\sigma(T) = \max_{1 \leq |t| \leq T} |\zeta(\sigma + it)|.$$

(The restriction $|t| \geq 1$ is only added to avoid problems for the case $\sigma = 1$.) The following results hold for large T.

(a) $Z_\sigma(T) \to \zeta(\sigma)$ for $\sigma > 1$.

(b) For $\sigma = 1$, unconditionally it is known that $Z_1(T) = O((\log T)^{\frac{2}{3}})$, while on RH

$$Z_1(T) \lesssim 2e^\gamma \log\log T.$$

On the other hand, Granville and Soundararajan [13] proved that

$$Z_1(T) \geq e^{\gamma}(\log\log T + \log\log\log T - \log\log\log\log T),$$

for some arbitrarily large T. They further conjectured that it equals the above with an $O(1)$ term instead of the quadruple log-term.

(c) For $\frac{1}{2} < \sigma < 1$, unconditionally one has $Z_\sigma(T) = O(T^a)$ for various $a > 0$, while on RH

$$\log Z_\sigma(T) \leq A \frac{(\log T)^{2-2\sigma}}{(1-\sigma)\log\log T}$$

for some constant A. Montgomery [27] showed that

$$\log Z_\sigma(T) \geq \frac{\sqrt{\sigma - 1/2}}{20} \frac{(\log T)^{1-\sigma}}{(\log\log T)^{\sigma}}$$

and, using a heuristic argument, he conjectured that this is (apart from the constant) the correct order of $\log Z_\sigma(T)$. In a recent paper (see [24]), Lamzouri suggests that

$$\log Z_\sigma(T) \sim C(\sigma) \frac{(\log T)^{1-\sigma}}{(\log\log T)^{\sigma}}$$

with $C(\sigma) = G_1(\sigma)^{\sigma}\sigma^{-2\sigma}(1-\sigma)^{\sigma-1}$, where

$$G_1(x) = \int_0^{\infty} \log\left(\sum_{n=0}^{\infty} \frac{(u/2)^{2n}}{(n!)^2}\right) \frac{du}{u^{1+1/x}}.$$

(d) For $\sigma = \frac{1}{2}$ unconditionally one has $Z_{\frac{1}{2}}(T) = O(T^{\frac{32}{205}}(\log T)^c) = O(T^{0.156..})$ (see [19]), while on RH

$$\log Z_{\frac{1}{2}}(T) \leq A \frac{\log T}{\log\log T}$$

for some constant A. On the other hand, it is known that

$$\log Z_{\frac{1}{2}}(T) \geq c\sqrt{\frac{\log T}{\log\log T}}$$

(see [2], [31]). Using a heuristic argument, Farmer et al. ([8]) conjectured that

$$\log Z_{\frac{1}{2}}(T) \sim \sqrt{\frac{1}{2}\log T \log\log T}.$$

(e) For $\sigma < \frac{1}{2}$, the functional equation for $\zeta(s)$ reduces the problem to the case $\sigma > \frac{1}{2}$. So, for example,

$$Z_\sigma(T) \sim \zeta(1-\sigma)\left(\frac{T}{2\pi}\right)^{\frac{1}{2}-\sigma} \quad \text{for } \sigma < 0.$$

Mean values. For $\sigma \in (\frac{1}{2}, 1)$, the mean-value formula

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |\zeta(\sigma - it)|^2 \, dt = \sum_{n=1}^{\infty} \frac{1}{n^{2\sigma}} = \zeta(2\sigma),$$

is well-known (see [33]). Furthermore,

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |\zeta(\sigma - it)|^4 \, dt = \sum_{n=1}^{\infty} \frac{d(n)^2}{n^{2\sigma}} = \frac{\zeta(2\sigma)^4}{\zeta(4\sigma)}.$$

For higher powers, however, present knowledge is very patchy. It is expected that the above formulas extend to all higher moments, i.e.,

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |\zeta(\sigma - it)|^{2k} \, dt = \sum_{n=1}^{\infty} \frac{d_k(n)^2}{n^{2\sigma}}.$$

This is equivalent to the Lindelöf Hypothesis.

Examples.

(a) The mean values for ζ_σ and ζ_σ^2 imply that $\zeta_\sigma, \zeta_\sigma^2 \in B_{\mathbb{N}}^2$ for $\sigma \in (\frac{1}{2}, 1)$. Note that this also implies $|\zeta_\sigma|^2 \in B_{\mathbb{Q}^+}^2$.

For higher powers, however, only partial results are known. For example, it is known that $\zeta_\sigma^k \in B_{\mathbb{N}}^2$ if $\sigma \in (1 - \frac{1}{k}, 1)$. Slightly better bounds are available, especially for particular values of k, but it is expected that much more holds, namely: $\zeta_\sigma^k \in B_{\mathbb{N}}^2$ for every $k \in \mathbb{N}$ and all $\sigma \in (\frac{1}{2}, 1)$. This is (equivalent to) the Lindelöf Hypothesis.

(b) Let $g(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}$ and $h(s) = \sum_{n=1}^{\infty} \frac{b_n}{n^s}$ be two Dirichlet series which converge absolutely for $\Re s > \sigma_0$ and $\Re s > \sigma_1$, respectively. Let $\alpha > \sigma_0$ and $\beta > \sigma_1$ and put $f(t) = g(\alpha - it)h(\beta + it)$. Then $f \in W_{\mathbb{Q}^+}$ with

$$\hat{f}(q) = \frac{1}{m^\alpha n^\beta} \sum_{d=1}^{\infty} \frac{a_{md} b_{nd}}{d^{\alpha+\beta}} \qquad \text{for } q = \tfrac{m}{n} \text{ with } (m,n) = 1.$$

We can prove this by multiplying out the series for $g(\alpha - it)$ and $h(\beta + it)$. We have

$$\begin{aligned} f(t) &= \sum_{m=1}^{\infty} \frac{a_m}{m^{\alpha - it}} \sum_{n=1}^{\infty} \frac{b_n}{n^{\beta + it}} = \sum_{m,n\ge 1} \frac{a_m b_n}{m^\alpha n^\beta} \Big(\frac{m}{n}\Big)^{it} \\ &= \sum_{d=1}^{\infty} \sum_{\substack{m,n\ge 1 \\ (m,n)=d}} \frac{a_m b_n}{m^\alpha n^\beta} \Big(\frac{m}{n}\Big)^{it} = \sum_{d=1}^{\infty} \frac{1}{d^{\alpha+\beta}} \sum_{\substack{m,n\ge 1 \\ (m,n)=1}} \frac{a_{md} b_{nd}}{m^\alpha n^\beta} \Big(\frac{m}{n}\Big)^{it} \\ &= \sum_{\substack{m,n\ge 1 \\ (m,n)=1}} \frac{1}{m^\alpha n^\beta} \Big(\sum_{d=1}^{\infty} \frac{a_{md} b_{nd}}{d^{\alpha+\beta}}\Big) \Big(\frac{m}{n}\Big)^{it}. \end{aligned}$$

Further Properties of $W_{\mathbb{Q}^+}$ and $W_{\mathbb{N}}$. Notation. For a unital Banach algebra $\mathcal{A}$, denote by $G\mathcal{A}$ the group of invertible elements of $\mathcal{A}$, and by $G_0\mathcal{A}$, the connected component of $G\mathcal{A}$ which contains the identity. If $\mathcal{A}$ is commutative, then

$$a \in G_0\mathcal{A} \Longleftrightarrow a = e^b \quad \text{for some } b \in \mathcal{A}.$$

(a) Wiener's Theorem for $W_{\mathbb{Q}^+}$ and $W_{\mathbb{N}}$ (see [15], Theorems 1 and 2):

Let $f \in W_{\mathbb{Q}^+}$. Then $1/f \in W_{\mathbb{Q}^+}$ if and only if $\inf_{\mathbb{R}} |f| > 0$; i.e., $GW_{\mathbb{Q}^+} = \{f \in W_{\mathbb{Q}^+} : \inf_{\mathbb{R}} |f| > 0\}$.

Let $f(t) = \sum_{n=1}^{\infty} a_n n^{it} \in W_{\mathbb{N}}$ and put $F(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}$ for $\Re s \geq 0$. Then $1/f \in W_{\mathbb{N}}$ if and only if there exists $\delta > 0$ such that $|F(s)| \geq \delta$ for all $\Re s \geq 0$.

The example $f(t) = 2^{it}$ shows that the condition $|f(t)| \geq \delta > 0$ for all $t \in \mathbb{R}$ is not sufficient for $1/f \in W_{\mathbb{N}}$.

(b) Let $f \in GW_{\mathbb{Q}^+}$. Then the *average winding number*[1] $w(f)$, defined by

$$w(f) = \lim_{T \to \infty} \frac{\arg f(T) - \arg f(-T)}{2T}$$

exists, and $w(f) = \log q$ for some $q \in \mathbb{Q}^+$ (see [21], Theorem 27).

It is easy to see that (i) $w(fg) = w(f) + w(g)$, and (ii) $w(\chi_q) = \log q$.

(c) $G_0 W_{\mathbb{N}} = GW_{\mathbb{N}}$; i.e. for $f \in W_{\mathbb{N}}$, we have $1/f \in W_{\mathbb{N}}$ if and only if $f = e^g$ for some $g \in W_{\mathbb{N}}$

3.4 Factorization and invertibility of multiplicative Toeplitz operators

The analogue of the factorization $T(abc) = T(a)T(b)T(c)$ for $a \in \overline{H^\infty}, b \in L^\infty, c \in H^\infty$ for Toeplitz matrices holds for Multiplicative Toeplitz matrices.

Let $\overline{W_{\mathbb{N}}}$ denote the set of absolutely convergent series $\sum_{n=1}^{\infty} a_n n^{-it}$.

Theorem 3.3. *Let $f \in \overline{W_{\mathbb{N}}}$, $g \in W_{\mathbb{Q}^+}$, and $h \in W_{\mathbb{N}}$. Then*

$$M(fgh) = M(f)M(g)M(h).$$

Proof. We show the matrix entries agree. By Proposition 3.1, $fgh \in W_{\mathbb{Q}^+}$ and

$$\begin{aligned}\Big(M(f)M(g)M(h)\Big)_{ij} &= \sum_{k,l \geq 1} M(f)_{ik} M(g)_{kl} M(h)_{lj} \\ &= \sum_{\substack{k,\,l \,\geq\, 1 \\ i\,|k \text{ and } j\,|l}} \hat{f}\Big(\frac{i}{k}\Big)\hat{g}\Big(\frac{k}{l}\Big)\hat{h}\Big(\frac{l}{j}\Big) \\ &= \sum_{m,n \geq 1} \hat{f}\Big(\frac{1}{m}\Big)\hat{g}\Big(\frac{mi}{nj}\Big)\hat{h}\Big(\frac{n}{1}\Big).\end{aligned}$$

[1] Also known as *mean motion*.

On the other hand, since all $\mathbb{Q}^+$-series converge absolutely we have

$$\begin{aligned} f(t)g(t)h(t) &= \sum_{q_1,q_2,q_3\in\mathbb{Q}^+} \hat{f}(q_1)\hat{g}(q_2)\hat{h}(q_3)(q_1q_2q_3)^{it} \\ &= \sum_{q\in\mathbb{Q}^+}\Big(\sum_{q_1,q_3}\hat{f}(q_1)\hat{g}\Big(\frac{q}{q_1q_3}\Big)\hat{h}(q_3)\Big)q^{it} \\ &= \sum_{q\in\mathbb{Q}^+}\widehat{fgh}(q)q^{it}. \end{aligned}$$

Hence

$$M(fgh)_{ij} = \widehat{fgh}\Big(\frac{i}{j}\Big) = \sum_{q_1,q_3}\hat{f}(q_1)\hat{g}\Big(\frac{i}{jq_1q_3}\Big)\hat{h}(q_3) = \Big(M(f)M(g)M(h)\Big)_{ij},$$

since $1/q_1$ and q_3 must range over $\mathbb{N}$. □

In view of Theorem 3.3, it is of interest to know when a given $f \in W_{\mathbb{Q}^+}$ factorises as $f = gh$ with $g \in \overline{W_{\mathbb{N}}}$ and $h \in W_{\mathbb{N}}$. For then $M(f) = M(g)M(h)$ and the invertibility of $M(f)$ follows from knowing when $M(g)$ and $M(h)$ are invertible. Thus, if h is invertible in $W_{\mathbb{N}}$, then

$$M(h)M(h^{-1}) = M(hh^{-1}) = M(1) = I = M(1) = M(h^{-1}h) = M(h^{-1})M(h),$$

so that $M(h)^{-1} = M(h^{-1})$. Similarly, if g is invertible in $\overline{W_{\mathbb{N}}}$, then $M(g)^{-1} = M(g^{-1})$. It follows then that $M(f)^{-1} = M(h)^{-1}M(g)^{-1}$.

Let $\mathcal{F}W_{\mathbb{Q}^+}$ denote the set of functions in $W_{\mathbb{Q}^+}$ which factorise as

$$f = f_-\chi_q f_+ \tag{3.9}$$

where $f_- \in G\overline{W_{\mathbb{N}}}$, $f_+ \in GW_{\mathbb{N}}$, and $q \in \mathbb{Q}^+$.

Note that with f as above, then $1/f = f_-^{-1}\chi_{(1/q)}f_+^{-1}$, so $1/f \in \mathcal{F}W_{\mathbb{Q}^+}$. In particular, $\mathcal{F}W_{\mathbb{Q}^+} \subset GW_{\mathbb{Q}^+}$. Note that $M(\chi_q)$ is invertible if and only if $q = 1$.

Theorem 3.4. *Let $f \in \mathcal{F}W_{\mathbb{Q}^+}$. Then $M(f)$ is invertible if and only if $w(f) = 0$. If this is the case, then $M(f)^{-1} = M(f_+^{-1})M(f_-^{-1})$, with $f_\pm$ as in (3.9).*

Proof. Write $f = f_-\chi_q f_+$ as in (3.9). Then $M(f) = M(f_-)M(\chi_q)M(f_+)$. Now $M(f_-)$ and $M(f_+)$ are invertible, with inverses $M(f_-^{-1})$ and $M(f_+^{-1})$, respectively. Thus $M(f)$ is invertible if and only if $M(\chi_q)$ is invertible. But this happens if and only if $q = 1$.

Since $w(f) = w(f_-) + w(\chi_q) + w(f_+) = w(\chi_q) = \log q$, we see that $M(f)$ is invertible if and only if $w(f) = 0$.

Now suppose $w(f) = 0$. Then the above gives

$$M(f)^{-1} = \Big(M(f_-)M(f_+)\Big)^{-1} = M(f_+)^{-1}M(f_-)^{-1} = M(f_+^{-1})M(f_-^{-1}). \quad \square$$

Multiplicative coefficients and Euler products.

Definition.

(a) A function $a\colon \mathbb{Q}^+ \to \mathbb{C}$ is *multiplicative* if $a(1) = 1$ and

$$a(p_1^{a_1} \cdots p_k^{a_k}) = a(p_1^{a_1}) \cdots a(p_k^{a_k}),$$

for all distinct primes p_i and all $a_i \in \mathbb{Z}$. We say a is *completely multiplicative* if, in addition to the above,

$$a(p^k) = a(p)^k \quad \text{and} \quad a(p^{-k}) = a(p^{-1})^k,$$

for all primes p and $k \in \mathbb{N}$.

(b) For a subset S of $\mathcal{F}$, let $\mathcal{M}\mathcal{F}$ denote the set of $f \in S$ for which $\hat{f}(\cdot)$ is multiplicative.

(c) Let $f \in \mathcal{F}$ and p prime. Suppose that $\sum_{k\in\mathbb{Z}} |\hat{f}(p^k)|$ converges. Then define

$$f_p = \sum_{k\in\mathbb{Z}} \hat{f}(p^k)\chi_{p^k}.$$

Note that f_p is periodic with period $\frac{2\pi}{\log p}$. Define $f_p^\sharp\colon \mathbb{T} \to \mathbb{C}$ by

$$f_p^\sharp(e^{i\theta}) = f_p(\theta/\log p) = \sum_{k\in\mathbb{Z}} \hat{f}(p^k)e^{ki\theta} \qquad \text{for } 0 \le \theta < 2\pi.$$

Further, denote by $W_{\mathbb{Q}^+,p}$ the set of those $f \in W_{\mathbb{Q}^+}$ whose $\mathbb{Q}^+$-coefficients are supported on $\{p^k : k \in \mathbb{Z}\}$. (Thus $f_p \in W_{\mathbb{Q}^+,p}$ by definition.)

Note that, for fixed p, there is a one-to-one correspondence between $W_{\mathbb{Q}^+,p}$ and $W(\mathbb{T})$ via the mapping $^\sharp$.

In [16], the Euler product formulas

$$f = \sum_{q\in\mathbb{Q}^+} \hat{f}(q)\chi_q = \prod_p f_p \quad \text{and} \quad M(f) = \prod_p M(f_p),$$

were shown to hold whenever $f \in \mathcal{M}W_{\mathbb{Q}^+}$.

The analogue of the Wiener–Hopf factorization holds for $\mathcal{M}W_{\mathbb{Q}^+}$-functions without zeros.

Theorem 3.5. *Let $f \in \mathcal{M}W_{\mathbb{Q}^+}$ such that f has no zeros. Then $f \in \mathcal{F}W_{\mathbb{Q}^+}$.*

Proof. We have $f = \prod_p f_p$, where $f_p \in W_{\mathbb{Q}^+,p}$ and each is non-zero. Hence $f_p^\sharp \in W(\mathbb{T})$ and has no zeros. Let $k_p = \operatorname{wind}(f_p^\sharp, 0)$. Note that $k_p = 0$ for all sufficiently large p, since

$$|f_p^\sharp(t)-1| \le \sum_{m\neq 0} |\hat{f}(p^m)| \le \sum_{|q|\ge p} |\hat{f}(q)| \longrightarrow 0 \quad \text{as } p\to\infty.$$

Let $q=\prod_p p^{k_p}$ (a finite product).

By Section 2.1(iii), we have

$$f_p^\sharp = \chi^\sharp_{p^{k_p}} e^{g_p^\sharp},$$

for some $g_p^\sharp \in W(\mathbb{T})$. Hence

$$f_p = \chi_{p^{k_p}} e^{g_p},$$

with $g_p \in W_{\mathbb{Q}^+,p}$. Thus for P so large that $k_p=0$ for $p>P$, we have

$$\prod_{p\le P} f_p = \chi_q \exp\biggl\{\sum_{p\le P} g_p\biggr\}.$$

Now $f_p(t)\to 1$ as $p\to\infty$ uniformly in t, so can choose g_p so that $g_p(t)\to 0$ as $p\to\infty$ (uniformly in t). Hence, for all sufficiently large p (and all t), $|f_p-1| = |e^{g_p}-1| \le \frac{1}{2}|g_p|$, so that

$$|g_p| \le 2|f_p-1| \le 2\sum_{m\neq 0} |\hat{f}(p^m)|.$$

Let $g^{(n)}=\sum_{p\le n} g_p$. Then $\{g^{(n)}\}$ is a Cauchy sequence in $W_{\mathbb{Q}^+}$: for $n>m$

$$|g^{(n)}-g^{(m)}| \le \sum_{m<p\le n} |g_p| \le 2\sum_{m<p\le n}\sum_{m\neq 0} |\hat{f}(p^m)| \le 2\sum_{r>m} |\hat{f}(k)| \longrightarrow 0$$

as $m\to\infty$. Thus $g^{(n)}\to g\in W_{\mathbb{Q}^+}$. But each $g^{(n)}$ is of the form h_n+k_n with $h_n\in W_{\mathbb{N}}$ and $k_n\in\overline{W_{\mathbb{N}}}$ (since $g_p\in W_{\mathbb{Q}^+,p}$). Thus $g=h+k$ with $h\in W_{\mathbb{N}}$, $k\in\overline{W_{\mathbb{N}}}$. It follows that $f=\chi_q e^h e^k$, which is of the required form. □

Note that, as such,

$$w(f)=\sum_p w(\chi_p^{k_p}) = \sum_p k_p w(\chi_p) = \sum_p \operatorname{wind}(f_p^\sharp,0)\log p,$$

where the sum is finite.

Corollary 3.6. *Let $f\in\mathcal{M}W_{\mathbb{Q}^+}$ such that f has no zeros and $w(f)=0$. Then $M(f)$ is invertible.*

Example. $M(\zeta_\alpha)$ is invertible for $\alpha>1$ with $M(\zeta_\alpha)^{-1}=M(1/\zeta_\alpha)$.

4 Unbounded multiplicative Toeplitz operators and matrices

The last section has, in various ways, been a relatively straightforward extension of the theory of bounded Toeplitz operators to the multiplicative setting. The theory of *unbounded* Toeplitz operators is rather less satisfactory and not easy to generalise. Our particular concern is in fact the operator $M(\zeta_\alpha)$ for $\alpha \leq 1$, since we expect a connection with the Riemann zeta function. The hope is that if a satisfactory theory is developed for this case, it can be generalised to other symbols.

In this section, we shall therefore concentrate on the particular case when $f(n) = n^{-\alpha}$, when the symbol is $\zeta(\alpha - it)$. Throughout we write φ_α (equivalently, $M(\zeta_\alpha)$) for φ_f.

From Section 3 we see that for $\alpha \leq 1$, φ_α is unbounded. It is interesting to see to what extent properties of φ_α are related to properties of ζ_α. The above theory is only valid for absolutely convergent Dirichlet series, when the symbols are bounded. But $\zeta(\alpha - it)$ is unbounded for $\alpha \leq 1$.

How to measure unboundedness? We shall investigate two different measures. The first case can be considered as restricting the range, while in the second case we shall restrict the domain.

4.1 First measure – the function $\Phi_\alpha(N)$

With b_n defined by $a = (a_n) \overset{\varphi_\alpha}{\mapsto} (b_n)$, i.e., $b_n = \sum_{d|n} d^{-\alpha} a_{n/d}$, let

$$\Phi_\alpha(N) = \sup_{\|a\|=1} \left(\sum_{n=1}^{N} |b_n|^2 \right)^{1/2}.$$

Theorem 4.1. *We have the following asymptotic formulae for large N:*

$$\Phi_1(N) = e^\gamma \log\log N + O(1) \qquad (\alpha = 1)$$

$$\log \Phi_\alpha(N) \asymp \frac{(\log N)^{1-\alpha}}{\log\log N} \qquad \left(\frac{1}{2} < \alpha < 1\right)$$

$$\log \Phi_{\frac{1}{2}}(N) \sim \left(\frac{\log N}{\log\log N}\right)^{\frac{1}{2}}. \qquad \left(\alpha = \frac{1}{2}\right)$$

Sketch of proof. (*For the proof see* [17]). We start with upper bounds.

First we note that for any positive arithmetical function $g(n)$,

$$\Phi_\alpha(N)^2 \leq \left(\sum_{n \leq N} \frac{g(n)}{n^\alpha} \right) \cdot \left(\max_{n \leq N} \sum_{d|n} \frac{1}{g(d) d^\alpha} \right). \tag{3.10}$$

This is because

$$|b_n|^2 = \left| \sum_{d|n} \frac{1}{\sqrt{g(d)} d^{\frac{\alpha}{2}}} \cdot \frac{\sqrt{g(d)} a_{n/d}}{d^{\frac{\alpha}{2}}} \right|^2 \leq \left(\sum_{d|n} \frac{1}{g(d) d^\alpha} \right) \cdot \left(\sum_{d|n} \frac{g(d) |a_{n/d}|^2}{d^\alpha} \right),$$

by Cauchy–Schwarz. Writing $G(n) = \sum_{d|n} g(d)^{-1}d^{-\alpha}$, we have

$$\sum_{n\le N}|b_n|^2 \le \sum_{n\le N} G(n)\sum_{d|n}\frac{g(d)|a_{n/d}|^2}{d^\alpha} \le \max_{n\le N} G(n)\sum_{d\le N}\frac{g(d)}{d^\alpha}\sum_{n\le N/d}|a_n|^2.$$

Taking $\|a\|_2 = 1$ yields (3.10). The idea is to choose g appropriately, so that the RHS of (3.10) is as small as possible.

For $\frac{1}{2} < \alpha \le 1$, choose $g(n)$ to be the following multiplicative function: for a prime power p^k let

$$g(p^k) = \begin{cases} 1, & \text{if } p^k \le M, \\ (\frac{M}{p^k})^\beta, & \text{if } p^k > M. \end{cases}$$

Here $M = (2\alpha - 1)\log N$ and β is a constant satisfying $1 - \alpha < \beta < \alpha$. Note that $g(p^k) \le g(p)$ for every $k \in \mathbb{N}$ and p prime.

We estimate the expressions in (3.10) separately. First

$$\sum_{n\le N}\frac{g(n)}{n^\alpha} \le \prod_p\Bigl(1+\sum_{k=1}^{\infty}\frac{g(p^k)}{p^{k\alpha}}\Bigr) \le \prod_p\Bigl(1+\frac{g(p)}{p^\alpha-1}\Bigr) \le \exp\Bigl\{\sum_p\frac{g(p)}{p^\alpha-1}\Bigr\} \tag{3.11}$$

and, after some manipulations using the Prime Number Theorem, one finds for the case $\alpha < 1$

$$\log\sum_{n\le N}\frac{g(n)}{n^\alpha} \lesssim \frac{\beta M^{1-\alpha}}{(1-\alpha)(\alpha+\beta-1)\log M}. \tag{3.12}$$

Now consider $G(n)$, which is multiplicative because $g(n)$ is. At the prime powers we have

$$\begin{aligned} G(p^k) = \sum_{r=0}^{k}\frac{1}{p^{\alpha r}g(p^r)} &= \sum_{\substack{r\ge 0\\ p^r\le M}}\frac{1}{p^{\alpha r}} + \frac{1}{M^\beta}\sum_{\substack{r\le k\\ p^r > M}}\frac{1}{p^{(\alpha-\beta)r}} \\ &\le 1 + \frac{1}{p^\alpha - 1} + \frac{1}{M^\alpha(1-p^{\beta-\alpha})}. \end{aligned}$$

(Here we require $\beta < \alpha$.) Note that this is independent of k. It follows that

$$G(n) \le \exp\Bigl\{\sum_{p|n}\frac{1}{p^\alpha-1} + \frac{1}{M^\alpha}\sum_{p|n}\frac{1}{1-p^{\beta-\alpha}}\Bigr\}.$$

The right-hand side is maximised when n is as large as possible (i.e. N) and N is of the form $N = 2.3\ldots P$. For such a choice, $\log N = \theta(P) \sim P$, so that (using the prime number theorem)

$$\log\max_{n\le N} G(n) \lesssim \sum_{p\le P}\frac{1}{p^\alpha-1} + \frac{1}{M^\alpha}\sum_{p\le P}1 \sim \frac{(\log N)^{1-\alpha}}{(1-\alpha)\log\log N} + \frac{\log N}{M^\alpha\log\log N}. \tag{3.13}$$

From (3.12) and (3.13) it follows that M should be of order $\log N$ for optimality. So taking $M = \lambda \log N$ (with $\lambda > 0$), (3.10), (3.12) and (3.13) then imply

$$\log \Phi_\alpha(N) \lesssim \left(\frac{\beta\lambda^{1-\alpha}}{2(1-\alpha)(\alpha+\beta-1)} + \frac{1}{2(1-\alpha)} + \frac{1}{2\lambda^\alpha} \right) \frac{(\log N)^{1-\alpha}}{\log\log N}$$

for every $\beta \in (1-\alpha, \alpha)$ and $\lambda > 0$. Since $\frac{\beta}{\alpha+\beta-1}$ decreases with β, the optimal choice is to take β arbitrarily close to α. Hence we require $\inf_{\lambda>0} h(\lambda)$, where

$$h(\lambda) = \frac{\alpha\lambda^{1-\alpha}}{(1-\alpha)(2\alpha-1)} + \frac{1}{(1-\alpha)} + \frac{1}{\lambda^\alpha}.$$

Since $h'(\lambda) = \frac{\alpha}{\lambda^{\alpha+1}}(\frac{\lambda}{2\alpha-1} - 1)$, we see that the optimal choice is indeed $\lambda = 2\alpha - 1$. Substituting this value of λ gives

$$\log \Phi_\alpha(N) \lesssim \frac{(1 + (2\alpha-1)^{-\alpha})}{2(1-\alpha)} \frac{(\log N)^{1-\alpha}}{\log\log N}.$$

For $\alpha = 1$, we use the same function $g(n)$ as before (though with possibly different values of M and β). From (3.11) it follows that

$$\sum_{n\le N} \frac{g(n)}{n} \le \prod_{p\le M} \left(\frac{1}{1-\frac{1}{p}} \right) \cdot \prod_{p>M} \left(1 + \frac{M^\beta}{p^\beta(p-1)} \right).$$

By Mertens' Theorem, the first product is $e^\gamma \log M + O(1)$, while $M^\beta \sum_{p>M} p^{-1-\beta} = O(1/\log M)$, and so

$$\sum_{n\le N} \frac{g(n)}{n} \le \Big(e^\gamma \log M + O(1) \Big) \exp\{O(1/\log M)\} = e^\gamma \log M + O(1). \quad (3.14)$$

For the $G(n)$ term we have, as for the $\alpha < 1$ case,

$$G(p^k) \le \frac{1}{1-\frac{1}{p}} + \frac{1}{M(1-p^{\beta-1})}.$$

Thus, with $N = 2.3 \ldots P$,

$$\begin{aligned} G(N) &\le \prod_{p\le P} \left(\frac{1}{1-\frac{1}{p}} \right) \left(1 + \frac{1-1/p}{M(1-p^{\beta-1})} \right) \\ &= \Big(e^\gamma \log P + O(1) \Big) \Big(1 + O\Big(\frac{P}{M \log P} \Big) \Big). \end{aligned}$$

Taking $M = \log N$ and noting that $P \sim \log N$, the right-hand side is $e^\gamma \log\log N + O(1)$. Combining with (3.14) shows that

$$\Phi_1(N) \le e^\gamma \log\log N + O(1).$$

The case $\alpha = \frac{1}{2}$. The function g as chosen for $\alpha \in (\frac{1}{2}, 1]$ is not suitable for an upper bound as we would require $\frac{1}{2} < \beta < \frac{1}{2}$! Instead we take g to be the multiplicative function following: for a prime power p^k, let

$$g(p^k) = \min\Big\{1, \Big(\frac{M}{p^k(\log p)^2}\Big)^{\frac{1}{2}}\Big\}.$$

Here $M > 0$ is independent of p and k and will be determined later. Thus $g(p^k) = 1$ if and only if $p^k(\log p)^2 \le M$. Note that $g(p^k) \le g(p) \le 1$ for all $k \ge 1$ and all primes p. Thus (3.11) holds with $\alpha = \frac{1}{2}$ and (using the prime number theorem)

$$\log \sum_{n \le N} \frac{g(n)}{\sqrt{n}} \lesssim \sum_{p \lesssim \frac{M}{(\log M)^2}} \frac{1}{\sqrt{p} - 1} + \sqrt{M} \sum_{p \gtrsim \frac{M}{(\log M)^2}} \frac{1}{p \log p} \sim \frac{\sqrt{M}}{\log M}. \tag{3.15}$$

(The first sum is of order $\sqrt{M}/(\log M)^2$ and the main contribution comes from the second term.)

Regarding $G(n)$, this time we have[2]

$$G(n) = \prod_{p^k \| n} G(p^k) \le \prod_{p^k \| n} \Big(1 + \sum_{r=1}^{k} \frac{1}{p^{r/2}} + \frac{1}{\sqrt{M}} \sum_{r=1}^{k} \log p\Big),$$

so that

$$\log G(n) \le \sum_{p|n} \frac{1}{\sqrt{p} - 1} + \frac{1}{\sqrt{M}} \sum_{p^k \| n} k \log p \le \frac{\log n}{\sqrt{M}} + \sum_{p|n} \frac{1}{\sqrt{p} - 1}.$$

The right-hand side above is maximal when $n = N = 2.3 \dots P$, hence

$$\log \max_{n \le N} G(n) \lesssim \frac{\log N}{\sqrt{M}} + \sum_{p \le P} \frac{1}{\sqrt{p}} \sim \frac{\log N}{\sqrt{M}} + \frac{2\sqrt{\log N}}{\log \log N}.$$

Combining with (3.15), (3.10) then gives

$$\log \Phi_{\frac{1}{2}}(N) \lesssim \frac{\sqrt{M}}{2 \log M} + \frac{\log N}{2\sqrt{M}} + \frac{\sqrt{\log N}}{\log \log N}.$$

The optimal choice for M is easily seen to be $M = \log N \log \log N$, and this gives the upper bound in (iii).

Now we proceed to give lower bounds.

For a fixed $n \in \mathbb{N}$, let

$$a_d = \frac{1}{\sqrt{d(n)}} \qquad \text{if } d|n, \text{ and zero otherwise.}$$

[2] Here as usual, $p^k || n$ means $p^k | n$ but $p^{k+1} \not| n$.

Then $\|a\|_2 = 1$, while for $d|n$

$$b_d = \frac{1}{\sqrt{d(n)}}\sum_{c|d}\frac{1}{c^\alpha} = \frac{\sigma_{-\alpha}(d)}{\sqrt{d(n)}}.$$

Hence for $N \ge n$,

$$\sum_{k\le N}|b_k|^2 \ge \sum_{d|n} b_d^2 = \frac{1}{d(n)}\sum_{d|n}\sigma_{-\alpha}(d)^2 =: \eta_\alpha(n).$$

With this notation $\Phi_\alpha(N) \ge \max_{n\le N}\sqrt{\eta_\alpha(n)}$, and the lower bounds follow from the maximal order of $\eta_\alpha(n)$. For $\frac{1}{2} < \alpha < 1$ this can be found easily. Since $\eta_\alpha(p) = 1 + p^{-\alpha} + \frac{1}{2}p^{-2\alpha}$ for p prime, we find for $n = 2.3\ldots P$ (so that $\log n \sim P$) that

$$\begin{aligned}\eta_\alpha(n) &= \prod_{p\le P}\Big(1 + \frac{1}{p^\alpha} + O\Big(\frac{1}{p^{2\alpha}}\Big)\Big) = \exp\Big\{(1+o(1))\sum_{p\le P}\frac{1}{p^\alpha}\Big\}\\ &= \exp\left\{\frac{(1+o(1))P^{1-\alpha}}{(1-\alpha)\log P}\right\} = \exp\left\{\frac{(1+o(1))(\log n)^{1-\alpha}}{(1-\alpha)\log\log n}\right\}.\end{aligned}$$

Now, if t_k is the k^{th} number of the form $2.3\cdots P$ (i.e., $t_k = p_1\cdots p_k$), then $\log t_k \sim k\log k \sim \log t_{k+1}$. Hence for $t_k \le N < t_{k+1}$, $\log N \sim k\log k$. It follows that

$$\Phi_\alpha(N) \ge \sqrt{\eta_\alpha(t_k)} \ge \exp\left\{\frac{(1+o(1))(\log t_k)^{1-\alpha}}{2(1-\alpha)\log\log t_k}\right\} = \exp\left\{\frac{(1+o(1))(\log N)^{1-\alpha}}{2(1-\alpha)\log\log N}\right\}.$$

For $\alpha = 1$, we have to be a little subtler to obtain $\max_{n\le N}\sqrt{\eta_1(n)} = e^\gamma\log\log N + O(1)$. We omit the details, which can be found in [17].

For the case $\alpha = \frac{1}{2}$, the above choice doesn't give the correct order and we lose a power of $\log\log N$. Instead, we follow an idea of Soundararajan [31]. Let f be the multiplicative function supported on the squarefree numbers whose values at primes p is

$$f(p) = \begin{cases} (\frac{M}{p})^{1/2}\frac{1}{\log p}, & \text{for } M \le p \le R,\\ \quad 0, & \text{otherwise.}\end{cases}$$

Here $M = \log N(\log\log N)$ as before and $\log R = (\log M)^2$.

Now take $a_n = f(n)F(N)^{-1/2}$, where $F(N) = \sum_{n\le N} f(n)^2$ so that $\sum_{n\le N} a_n^2 = 1$. Then by Hölder's inequality

$$\begin{aligned}\Big(\sum_{n=1}^N b_n^2\Big)^{1/2} &\ge \sum_{n=1}^N a_n b_n = \frac{1}{F(N)}\sum_{n=1}^N \frac{f(n)}{\sqrt{n}}\sum_{d|n}\sqrt{d}\,f(d)\\ &= \frac{1}{F(N)}\sum_{n\le N}\frac{f(n)}{\sqrt{n}}\sum_{\substack{d\,\le\, N/n\\ (n,d)\,=\,1}} f(d)^2. \qquad (3.16)\end{aligned}$$

Now using ‘Rankin’s trick’[3] we have, for any $\beta > 0$

$$\begin{aligned}\sum_{n\le N}\frac{f(n)}{n^{1/2}}\sum_{\substack{d\le N/n\\(n,d)=1}} f(d)^2 &= \sum_{n\le N}\frac{f(n)}{n^{1/2}}\Bigg(\sum_{\substack{d\ge 1\\(n,d)=1}} f(d)^2 - \sum_{\substack{d>N/n\\(n,d)=1}} f(d)^2\Bigg)\\ &= \sum_{n\le N}\frac{f(n)}{n^{1/2}}\Bigg(\prod_{p\nmid n}\Big(1+f(p)^2\Big)+O\bigg(\Big(\frac{n}{N}\Big)^{\beta}\prod_{p\nmid n}\Big(1+p^{\beta}f(p)^2\Big)\bigg)\Bigg).\end{aligned} \tag{3.17}$$

The O-term in (3.17) is at most a constant times

$$\frac{1}{N^{\beta}}\sum_{n\le N} f(n)n^{\beta-1/2}\prod_{p\nmid n}\Big(1+p^{\beta}f(p)^2\Big)\le\frac{1}{N^{\beta}}\prod_{p}\Big(1+p^{\beta}f(p)^2+p^{\beta-1/2}f(p)\Big),$$

while the main term in (3.17) is (using Rankin’s trick again)

$$\prod_{p}\Big(1+f(p)^2+\frac{f(p)}{p^{1/2}}\Big)+O\bigg(\frac{1}{N^{\beta}}\prod_{p}\Big(1+f(p)^2+p^{\beta-1/2}f(p)\Big)\bigg).$$

Hence (3.16) implies

$$\begin{aligned}\bigg(\sum_{n=1}^{N} b_n^2\bigg)^{1/2} \quad\ge\quad & \frac{1}{F(N)}\Bigg(\prod_{p}\Big(1+f(p)^2+\frac{f(p)}{p^{1/2}}\Big)\\ &+O\bigg(\frac{1}{N^{\beta}}\prod_{p}\Big(1+p^{\beta}f(p)^2+p^{\beta-1/2}f(p)\Big)\bigg)\Bigg).\end{aligned}$$

The ratio of the O-term to the main term on the right is less than

$$\exp\bigg\{-\beta\log N+\sum_{M\le p\le R}(p^{\beta}-1)\Big(f(p)^2+\frac{f(p)}{p^{1/2}}\Big)\bigg\},$$

which equals

$$\exp\bigg\{-\beta\log N+\sum_{M\le p\le R}(p^{\beta}-1)\bigg(\frac{M}{p(\log p)^2}+\frac{M^{1/2}}{p(\log p)}\bigg)\bigg\}.$$

Take $\beta = (\log M)^{-3}$. The term involving $M^{1/2}$ is at most $(\log N)^{1/2+\varepsilon}$ for every $\varepsilon > 0$, while the remaining terms in the exponent are (by the

[3] If $c_n > 0$, then for any $\beta > 0$, $\sum_{n>x} c_n \le x^{-\beta}\sum_{n=1}^{\infty} n^{\beta}c_n$.

prime number theorem in the form $\pi(x) = \mathrm{li}(x) + O(x(\log x)^{-A})$ for all A)

$$\begin{aligned}
-\beta \log N &+ M \int_M^R \frac{t^\beta - 1}{t(\log t)^3}\, dt + O\Big(\frac{\log N}{(\log\log N)^A}\Big) \\
&= -\beta \log N + \beta M \int_M^R \frac{dt}{t(\log t)^2} + O\Big(\beta^2 M \int_M^R \frac{dt}{t \log t}\Big) \\
&\sim -\beta \frac{\log N \log\log\log N}{\log\log N},
\end{aligned}$$

after some calculations.

Finally, since $F(N) \leq \prod_p (1 + f(p)^2)$, this implies

$$\Phi_{1/2}(N) \geq \frac{1}{2} \prod_{M \leq p \leq R} \Big(1 + \frac{f(p)}{p^{1/2}(1 + f(p)^2)}\Big),$$

for all N sufficiently large. Hence

$$\log \Phi_{1/2}(N) \gtrsim M^{1/2} \sum_{M \leq p \leq R} \frac{1}{p(\log p)} \sim \Big(\frac{\log N}{\log\log N}\Big)^{1/2},$$

as required. □

Remark. The result for $\frac{1}{2} < \alpha < 1$ is

$$\frac{(\log N)^{1-\alpha}}{2(1-\alpha)\log\log N} \lesssim \log \Phi_\alpha(N) \lesssim \frac{(1 + (2\alpha - 1)^{-\alpha})}{2(1-\alpha)} \frac{(\log N)^{1-\alpha}}{\log\log N}.$$

It would be nice to obtain an asymptotic formula for $\log \Phi_\alpha(N)$. Indeed, it is possible to improve the lower bound at the cost of more work by using the method for the case $\alpha = \frac{1}{2}$, but we have not been able to obtain the same upper and lower limits.

4.2 Connections between $\Phi_\alpha(N)$ and the order of $|\zeta(\alpha + it)|$ The lower bounds obtained for $\Phi_\alpha(N)$ for $\frac{1}{2} < \alpha \leq 1$ can be used to obtain information regarding the maximum order of $\zeta(s)$ on the line $\Re s = \alpha$.

Proposition 4.2. *With $b_n = \sum_{d|n} d^{-\alpha} a_{n/d}$ we have, for any α,*

$$\sum_{n \leq N} |b_n|^2 = \sum_{m,n \leq N} \frac{a_m \overline{a_n} (m,n)^{2\alpha}}{m^\alpha n^\alpha} \sum_{k \leq \frac{N}{[m,n]}} \frac{1}{k^{2\alpha}}.$$

Proof. We have

$$|b_n|^2 = b_n \overline{b_n} = \frac{1}{n^{2\alpha}} \sum_{c|n, d|n} c^\alpha d^\alpha a_c \overline{a_d} = \frac{1}{n^{2\alpha}} \sum_{[c,d]|n} c^\alpha d^\alpha a_c \overline{a_d},$$

since $c|n, d|n$ if and only if $[c,d]|n$. Hence

$$\sum_{n\le N} |b_n|^2 = \sum_{c,d\le N} c^\alpha d^\alpha a_c \overline{a_d} \sum_{n\le N, [c,d]|n} \frac{1}{n^{2\alpha}} = \sum_{c,d\le N} \frac{c^\alpha d^\alpha a_c\overline{a_d}}{[c,d]^{2\alpha}} \sum_{k\le \frac{N}{[c,d]}} \frac{1}{k^{2\alpha}},$$

by writing $n = [c,d]k$. Since $(c,d)[c,d] = cd$, the result follows. □

It follows that if $a_n \ge 0$ for all n and $\alpha > \frac{1}{2}$, then

$$\sum_{n\le N} |b_n|^2 \le \zeta(2\alpha) \sum_{m,n\le N} \frac{a_m\overline{a_n}(m,n)^{2\alpha}}{(mn)^\alpha}. \tag{3.18}$$

Theorem 4.3. *Let $\frac{1}{2} < \alpha \le 1$ and let $a \in l^2$ with $\|a\|_2 = 1$. Let $A_N(t) = \sum_{n=1}^N a_n n^{it}$. Let $N \le T^\lambda$, where $0 < \lambda < \frac{2}{3}(\alpha - \frac{1}{2})$. Then for some $\eta > 0$,*

$$\frac{1}{T}\int_1^T |\zeta(\alpha+it)|^2 |A_N(t)|^2\,dt = \zeta(2\alpha)\sum_{m,n\le N}\frac{a_m\overline{a_n}(m,n)^{2\alpha}}{(mn)^\alpha} + O(T^{-\eta}). \tag{3.19}$$

Proof. We shall assume $\frac{1}{2} < \alpha < 1$, adjusting the proof for the case $\alpha = 1$ afterwards. For $\alpha \ne 1$, we can integrate from 0 to T since the error involved is at most $O(N/T) = O(T^{-\eta})$.

Starting from the approximation $\zeta(\alpha+it) = \sum_{n\le t} n^{-\alpha-it} + O(t^{-\alpha})$, we have

$$|\zeta(\alpha+it)|^2 = \left|\sum_{n\le t}\frac{1}{n^{\alpha+it}}\right|^2 + O(t^{1-2\alpha}).$$

Let $k, l \in \mathbb{N}$ such that $(k,l) = 1$. Let $M = \max\{k,l\} < T$. The above gives

$$\int_0^T |\zeta(\alpha+it)|^2\Big(\frac{k}{l}\Big)^{it}\,dt = \int_0^T \left|\sum_{n\le t}\frac{1}{n^{\alpha+it}}\right|^2\Big(\frac{k}{l}\Big)^{it}\,dt + O(T^{2-2\alpha}).$$

The integral on the right is

$$\int_0^T \sum_{m,n\le t}\frac{1}{(mn)^\alpha}\Big(\frac{km}{ln}\Big)^{it}\,dt = \sum_{m,n\le T}\frac{1}{(mn)^\alpha}\int_{\max\{m,n\}}^T \Big(\frac{km}{ln}\Big)^{it}\,dt.$$

The terms with $km = ln$ (which implies $m = rl, n = rk$ with r integral) contribute

$$\frac{1}{(kl)^\alpha}\sum_{r\le T/M}\frac{T - rM}{r^{2\alpha}} = \frac{\zeta(2\alpha)}{(kl)^\alpha}T + O\Big(\frac{M^{2\alpha-1}T^{2-2\alpha}}{(kl)^\alpha}\Big).$$

The remaining terms contribute at most

$$\begin{aligned}
2\sum_{\substack{m,n\le T\\ km\neq ln}} \frac{1}{(mn)^\alpha|\log\frac{km}{ln}|} &\le 2M^{2\alpha}\sum_{\substack{m,n\le T\\ km\neq ln}} \frac{1}{(kmln)^\alpha|\log\frac{km}{ln}|}\\
&\le 2M^{2\alpha}\sum_{\substack{m_1\le kT, n_1\le lT\\ m_1\neq n_1}} \frac{1}{(m_1n_1)^\alpha|\log\frac{m_1}{n_1}|}\\
&\le 2M^{2\alpha}\sum_{\substack{m_1,n_1\le MT\\ m_1\neq n_1}} \frac{1}{(m_1n_1)^\alpha|\log\frac{m_1}{n_1}|}\\
&= O(M^{2\alpha}(MT)^{2-2\alpha}\log(MT))\\
&= O(M^2T^{2-2\alpha}\log T),
\end{aligned}$$

using Lemma 7.2 from [33]. Hence

$$\int_0^T |\zeta(\alpha+it)|^2\Big(\frac{k}{l}\Big)^{it}\,dt = \frac{\zeta(2\alpha)}{(kl)^\alpha}T + O(M^2T^{2-2\alpha}\log T).$$

It follows that for any positive integers $m,n<T$,

$$\int_0^T |\zeta(\alpha+it)|^2\Big(\frac{m}{n}\Big)^{it}\,dt = \frac{\zeta(2\alpha)(m,n)^{2\alpha}}{(mn)^\alpha}T + O(\max\{m,n\}^2T^{2-2\alpha}\log T).$$

Thus, with $A_N(t)=\sum_{n=1}^N a_n n^{it}$,

$$\begin{aligned}
\int_0^T |\zeta(\alpha+it)|^2|A_N(t)|^2\,dt &= \sum_{m,n\le N} a_m\overline{a_n}\int_0^T |\zeta(\alpha+it)|^2\Big(\frac{m}{n}\Big)^{it}\,dt\\
&= \zeta(2\alpha)T\sum_{m,n\le N}\frac{a_m\overline{a_n}(m,n)^{2\alpha}}{(mn)^\alpha}\\
&\quad + O\Big(T^{2-2\alpha}\log T\sum_{m,n\le N}\max\{m,n\}^2|a_ma_n|\Big).
\end{aligned}$$

The sum in the O-term is at most $N^2(\sum_{n\le N}|a_n|)^2\le N^3$, using Cauchy–Schwarz. Hence

$$\frac{1}{T}\int_0^T |\zeta(\alpha+it)|^2|A_N(t)|^2\,dt = \zeta(2\alpha)\sum_{m,n\le N}\frac{a_m\overline{a_n}(m,n)^{2\alpha}}{(mn)^\alpha} + O\Big(\frac{N^3\log T}{T^{2\alpha-1}}\Big).$$

Since $N^3\le T^{3\lambda}$ and $3\lambda<2\alpha-1$, the error term is $O(T^{-\eta})$ for some $\eta>0$.

If $\alpha = 1$ we integrate from 1 to T instead and the O-term above will contain an extra $\log T$ factor, but this is still $O(T^{-\eta})$. □

We note that with more care the N^3 could be turned into an N^2, so that we can take $\lambda < \alpha - \frac{1}{2}$ in the theorem. This is however not too important for us.

Corollary 4.4. *Let $\frac{1}{2} < \alpha \leq 1$. Then for every $\varepsilon > 0$ and N sufficiently large,*

$$\max_{t \leq N} |\zeta(\alpha + it)| \geq \Phi_\alpha(N^{\frac{2}{3}(\alpha - \frac{1}{2}) - \varepsilon}) + O(N^{-\eta}) \tag{3.20}$$

for some $\eta > 0$.

Proof. Let $a_n \geq 0$ be such that $\|a\|_2 = 1$, and take $N = T^\lambda$ with $\lambda < \frac{2}{3}(\alpha - \frac{1}{2})$. By (3.18) and (3.19),

$$\begin{aligned}\sum_{n \leq N} |b_n|^2 &\leq \frac{1}{T}\int_0^T |\zeta(\alpha + it)|^2 |A_N(t)|^2\, dt + O(T^{-\eta}) \\ &\leq \max_{t \leq T} |\zeta(\alpha + it)|^2 \frac{1}{T}\int_0^T |A_N(t)|^2\, dt + O(T^{-\eta}) \\ &= \max_{t \leq T} |\zeta(\alpha + it)|^2 \sum_{n \leq N} |a_n|^2 (1 + O(N/T)) + O(T^{-\eta})\end{aligned}$$

using the Montgomery and Vaughan mean value theorem. The implied constants in the O-terms depend only on T and not on the sequence $\{a_n\}$. Taking the supremum over all such a, this gives

$$\Phi_\alpha(N)^2 = \sup_{\|a\|_2 = 1} \sum_{n \leq N} |b_n|^2 \leq \max_{t \leq T} |\zeta(\alpha + it)|^2 + O(T^{-\eta}),$$

for some $\eta > 0$, and (3.20) follows. □

In particular, this gives the (known) lower bounds

$$\max_{t \leq T} |\zeta(\alpha + it)| \geq \exp\left\{\frac{c(\log T)^{1-\alpha}}{\log\log T}\right\}$$

for $\frac{1}{2} < \alpha < 1$ and $\max_{t \leq T} |\zeta(1 + it)| \geq e^\gamma \log\log T + O(1)$ (obtained by Levinson in [25]).

Morever, we can say more about how often $|\zeta(\alpha + it)|$ is as large as this. For $A \in \mathbb{R}$ and $c > 0$, let

$$F_A(T) = \left\{t \in [1, T] : |\zeta(1 + it)| \geq e^\gamma \log\log T - A\right\}. \tag{3.21}$$

$$F_{\alpha,c}(T) = \left\{t \in [0, T] : |\zeta(\alpha + it)| \geq \exp\left\{\frac{c(\log T)^{1-\alpha}}{\log\log T}\right\}\right\}. \tag{3.21$'$}$$

We outline the argument in the case $\frac{1}{2} < \alpha < 1$. We have

$$\sum_{n \le N} |b_n|^2 \le \frac{1}{T}\left(\int_{F_{\alpha,c}(T)} + \int_{[0,T]\setminus F_{\alpha,c}(T)}\right)|\zeta(\alpha+it)|^2|A_N(t)|^2\,dt + O(T^{-\eta}). \tag{3.22}$$

The second integral on the right is at most

$$\exp\left\{\frac{2c(\log T)^{1-\alpha}}{\log\log T}\right\}\cdot\frac{1}{T}\int_0^T |A_N(t)|^2\,dt = O\left(\exp\left\{\frac{2c(\log T)^{1-\alpha}}{\log\log T}\right\}\right),$$

while, by choosing $a_n = d(N)^{-1/2}$ for $n|N$ and zero otherwise, the LHS of (3.22) is at least $\eta_\alpha(N)$. Every interval $[T^{\lambda/3}, T^{\lambda}]$ contains an N of the form $2.3.5\ldots P$. As such, $\eta_\alpha(N) \ge \exp\left\{\frac{c'(\log T)^{1-\alpha}}{\log\log T}\right\}$ for some $c' > 0$. Hence, for $2c < c'$,

$$\frac{1}{T}\int_{F_{\alpha,c}(T)} |\zeta(\alpha+it)|^2|A_N(t)|^2\,dt \ge \exp\left\{\frac{c'(\log T)^{1-\alpha}}{2\log\log T}\right\}.$$

We have $|\zeta(\alpha+it)| = O(T^{\nu})$ for some ν and $|A_N(t)|^2 \le d(N) = O(T^{\varepsilon})$, so

$$\frac{1}{T}\int_{F_{\alpha,c}(T)} |\zeta(\alpha+it)|^2|A_N(t)|^2\,dt \le T^{2\nu-1+\varepsilon}\mu(F_{\alpha,c}(T)).$$

Thus $\mu(F_{\alpha,c}(T)) \ge T^{1-2\nu-\varepsilon}$ for all c sufficiently small.

In particular, since $\nu < \frac{1-\alpha}{3}$, we have:

Theorem 4.5. *For all A sufficiently large (and positive),*

$$\mu(F_A(T)) \ge T\exp\left\{-a\frac{\log T}{\log\log T}\right\}$$

for some $a > 0$, and for all c sufficiently small, $\mu(F_{\alpha,c}(T)) \ge T^{(1+2\alpha)/3}$ for all sufficiently large T. Furthermore, under the Lindelöf Hypothesis, the exponent can be replaced by $1-\varepsilon$.

Addendum: There has been much recent activity in this field. Aistleitner [1], in essence restricting $(*)$ to N rows and columns (rather than just the first N) used an ingenious idea based on GCD-sums to improve these Ω-results for $|\zeta(\alpha+it)|$ and Theorem 3.5 to $\exp\left\{c\frac{(\log T)^{1-\alpha}}{(\log\log T)^{\alpha}}\right\}$ for $\frac{1}{2} < \alpha < 1$, thereby improving Montgomery's result. More recently still, Bondarenko and Seip [4] used a similar method for the case $\alpha = \frac{1}{2}$, obtaining

$$Z_{\frac{1}{2}}(T) \ge \exp\left\{c\sqrt{\frac{\log T\log\log\log T}{\log\log T}}\right\}$$

for any $c < \frac{1}{\sqrt{2}}$.

4.3 Second measure – φ_α on $\mathcal{M}^2$ and the function $M_\alpha(T)$ For $\alpha \le 1$, φ_α is unbounded on l^2 and so $\varphi_\alpha(l^2) \not\subset l^2$ (by the closed graph theorem[4] – see [32], p. 183). However, if we *restrict* the domain to $\mathcal{M}^2$, the set of multiplicative elements of l^2, we find that $\varphi(\mathcal{M}^2) \subset l^2$. More generally, if f is multiplicative then, as we shall see, $\varphi_f(\mathcal{M}^2) \subset l^2$ in many cases (and hence $\varphi_f(\mathcal{M}^2) \subset \mathcal{M}^2$).

Notation. Let $\mathcal{M}^2$ and $\mathcal{M}^2_{\mathrm{c}}$ denote the subsets of l^2 of multiplicative and completely multiplicative functions, respectively. Further, write $\mathcal{M}^2+$ for the non-negative members of $\mathcal{M}^2$, and similarly for $\mathcal{M}^2_{\mathrm{c}}+$.

Let $\mathcal{M}^2_0$ denote the set of $\mathcal{M}^2$ functions f for which $f * g \in \mathcal{M}^2$ whenever $g \in \mathcal{M}^2$; that is,

$$\mathcal{M}^2_0 = \{f \in \mathcal{M}^2 : g \in \mathcal{M}^2 \implies f * g \in \mathcal{M}^2\}.$$

Thus for $f \in \mathcal{M}^2_0$, $\varphi_f(\mathcal{M}^2) \subset \mathcal{M}^2$.

It was shown in [18] that $\mathcal{M}^2$ is not closed under Dirichlet convolution, so $\mathcal{M}^2_0 \neq \mathcal{M}^2$. A criterion for $f \in \mathcal{M}^2$ to be in $\mathcal{M}^2_0$ was given, namely:

Proposition 4.6. *Let $f \in \mathcal{M}^2$ be such that $\sum_{k=1}^\infty |f(p^k)|$ converges for every prime p and that $\sum_{k=1}^\infty |f(p^k)| \le A$ for some constant A independent of p. Then $f \in \mathcal{M}^2_0$.*

On the other hand, if $f \in \mathcal{M}^2$ with $f \ge 0$ and for some prime p_0, $f(p_0^k)$ decreases with k and $\sum_{k=1}^\infty f(p_0^k)$ diverges, then $f \notin \mathcal{M}^2_0$.

The proof is based on the following necessary and sufficient condition: *Let $f, g \in \mathcal{M}^2$ be non-negative. Then $f * g \in \mathcal{M}^2$ if and only if*

$$\sum_p \sum_{m,n\ge 1} \sum_{k=0}^\infty f(p^m)g(p^n)f(p^{m+k})g(p^{n+k}) \quad \text{converges.}$$

This can be proven in a direct manner.

Thus, in particular, $\mathcal{M}^2_{\mathrm{c}} \subset \mathcal{M}^2_0$. For $f \in \mathcal{M}^2_{\mathrm{c}}$ if and only if $|f(p)| < 1$ for all primes p and $\sum_p |f(p)|^2 < \infty$. Thus

$$\sum_{k=1}^\infty |f(p^k)| = \frac{|f(p)|}{1 - |f(p)|} \le A,$$

independent of p.

For example, $(n^{-\alpha}) \in \mathcal{M}^2_0$ for $\alpha > \frac{1}{2}$.

The "quasi-norm" $M_f(T)$. Let $f \in \mathcal{M}^2_0$. The discussion above shows that $\varphi_f(\mathcal{M}^2) \subset \mathcal{M}^2$ but, typically, φ_f is not 'bounded' on $\mathcal{M}^2$ (if $f \notin l^1$) in the sense

[4] Being a 'matrix' mapping, φ_α is necessarily a closed operator, and so $\varphi_\alpha(l^2) \subset l^2$ implies φ_α is bounded.

that[5] $\|\varphi_f a\|/\|a\|$ is not bounded by a constant for all $a \in \mathcal{M}^2$. A natural question is: how large can $\|\varphi_f a\|$ become as a function of $\|a\|$? It therefore makes sense to define, for $T \geq 1$,

$$M_f(T) = \sup_{\substack{a \in \mathcal{M}^2 \\ \|a\| = T}} \frac{\|\varphi_f a\|}{\|a\|}.$$

We shall consider only the case $f(n) = n^{-\alpha}$, although the result below can be extended without any significant changes to f completely multiplicative and such that $f|_{\mathbb{P}}$ is *regularly varying* of index $-\alpha$ with $\alpha > 1/2$ in the sense that there exists a regularly varying function $\tilde{f}$ (of index $-\alpha$) with $\tilde{f}(p) = f(p)$ for every prime p. We shall write $M_f(T) = M_\alpha(T)$ in this case.

Theorem 4.7. *As $T \to \infty$*

$$M_1(T) = e^\gamma(\log\log T + \log\log\log T + 2\log 2 - 1 + o(1)), \tag{3.23}$$

while for $\frac{1}{2} < \alpha < 1$,

$$\log M_\alpha(T) = \left(\frac{B(\frac{1}{\alpha}, 1 - \frac{1}{2\alpha})^\alpha}{(1-\alpha)2^\alpha} + o(1)\right)\frac{(\log T)^{1-\alpha}}{(\log\log T)^\alpha}. \tag{3.24}$$

Here $B(x,y)$ denotes the Beta function $\int_0^1 t^{x-1}(1-t)^{y-1}dt$.

Sketch of proof for the case $\frac{1}{2} < \alpha < 1$ (for the full proof see [18]*).* We consider first upper bounds. The supremum occurs for $a \geq 0$, which we now assume. Write $a = (a_n)$, $\varphi_\alpha a = b = (b_n)$. Define α_p and β_p for prime p by

$$\alpha_p = \sum_{k=1}^{\infty} a_{p^k}^2 \quad \text{and} \quad \beta_p = \sum_{k=1}^{\infty} b_{p^k}^2.$$

By the multiplicativity of a and b, $T^2 = \|a\|^2 = \prod_p(1+\alpha_p)$ and $\|b\|^2 = \prod_p(1+\beta_p)$. Thus

$$\frac{\|\varphi_\alpha a\|}{\|a\|} = \prod_p \sqrt{\frac{1+\beta_p}{1+\alpha_p}}.$$

Now for $k \geq 1$

$$b_{p^k} = \sum_{r=0}^{k} p^{-r\alpha} a_{p^{k-r}} = a_{p^k} + p^{-\alpha} b_{p^{k-1}},$$

whence

$$b_{p^k}^2 = a_{p^k}^2 + 2p^{-\alpha} a_{p^k} b_{p^{k-1}} + p^{-2\alpha} b_{p^{k-1}}^2.$$

[5] Here, and throughout this section $\|\cdot\| = \|\cdot\|_2$ is the l^2-norm.

Summing from $k = 1$ to ∞ and adding 1 to both sides gives

$$1 + \beta_p = 1 + \alpha_p + 2p^{-\alpha} \sum_{k=1}^{\infty} a_{p^k} b_{p^{k-1}} + p^{-2\alpha}(1 + \beta_p). \tag{3.25}$$

By Cauchy–Schwarz,

$$\sum_{k=1}^{\infty} a_{p^k} b_{p^{k-1}} \leq \left(\sum_{k=1}^{\infty} a_{p^k}^2 \sum_{k=1}^{\infty} b_{p^{k-1}}^2 \right)^{1/2} = \sqrt{\alpha_p(1 + \beta_p)},$$

so, on rearranging,

$$(1 + \beta_p) - \frac{2p^{-\alpha}\sqrt{\alpha_p(1 + \beta_p)}}{1 - p^{-2\alpha}} \leq \frac{1 + \alpha_p}{1 - p^{-2\alpha}}.$$

Completing the square we obtain

$$\left(\sqrt{1 + \beta_p} - \frac{p^{-\alpha}\sqrt{\alpha_p}}{1 - p^{-2\alpha}} \right)^2 \leq \frac{1 + \alpha_p}{(1 - p^{-2\alpha})^2}.$$

The term on the left inside the square is non-negative for every p since $1 + \beta_p \geq \frac{1+\alpha_p}{1-p^{-2\alpha}}$, which is greater than $\frac{p^{-2\alpha}\alpha_p}{(1-p^{-2\alpha})^2}$ for $\alpha > \frac{1}{2}$. Rearranging gives

$$\sqrt{\frac{1 + \beta_p}{1 + \alpha_p}} \leq \frac{1}{1 - p^{-2\alpha}} \left(1 + \frac{1}{p^{\alpha}} \sqrt{\frac{\alpha_p}{1 + \alpha_p}} \right).$$

Let $\gamma_p = \sqrt{\frac{\alpha_p}{1+\alpha_p}}$. Taking the product over all primes p gives

$$\frac{\|\varphi_\alpha a\|}{\|a\|} \leq \zeta(2\alpha) \prod_p \left(1 + \frac{\gamma_p}{p^{\alpha}} \right) \leq \zeta(2\alpha) \exp\left\{ \sum_p \frac{\gamma_p}{p^{\alpha}} \right\}. \tag{3.26}$$

Note that $0 \leq \gamma_p < 1$ and $\prod_p \frac{1}{1-\gamma_p^2} = T^2$. The idea is to show now that the main contribution to the above sum comes from the range $p \asymp \log T \log\log T$.

Let $\varepsilon > 0$ and put $P = \log T \log\log T$. We split up the sum on the right-hand side of (3.26) into $p \leq aP$, $aP < p \leq AP$, and $p > AP$ (for a small and A large). First,

$$\sum_{p \leq aP} p^{-\alpha}\gamma_p \leq \sum_{p \leq aP} p^{-\alpha} \sim \frac{a^{1-\alpha}P^{1-\alpha}}{(1-\alpha)\log P} < \varepsilon \frac{(\log T)^{1-\alpha}}{(\log\log T)^{\alpha}}, \tag{3.27}$$

for a sufficiently small. Next, using the fact that $\log T^2 = \log \prod_p \frac{1}{1-\gamma_p^2} \geq \sum_p \gamma_p^2$, we have

$$\sum_{p>AP} p^{-\alpha}\gamma_p \leq \Bigl(\sum_{p>AP} p^{-2\alpha} \sum_{p>AP} \gamma_p^2\Bigr)^{1/2} \lesssim \Bigl(\frac{2A^{1-2\alpha}P^{1-2\alpha}\log T}{(2\alpha-1)\log P}\Bigr)^{1/2}$$

$$\sim \frac{(\log T)^{1-\alpha}(\log\log T)^{-\alpha}}{A^{\alpha-1/2}\sqrt{\alpha-1/2}} < \varepsilon \frac{(\log T)^{1-\alpha}}{(\log\log T)^{\alpha}} \tag{3.28}$$

for A sufficiently large. This leaves the range $aP < p \leq AP$ and the problem therefore reduces to maximising

$$\sum_{aP<p\leq AP} \frac{\gamma_p}{p^\alpha}$$

subject to $0 \leq \gamma_p < 1$ and $\prod_p \frac{1}{1-\gamma_p^2} = T^2$. The maximum clearly occurs for γ_p decreasing (if $\gamma_{p'} > \gamma_p$ for primes $p < p'$, then the sum increases in value if we swap γ_p and $\gamma_{p'}$). Thus we may assume that γ_p is decreasing.

By interpolation we may write $\gamma_p = g(\frac{p}{P})$, where $g\colon (0,\infty) \to (0,1)$ is continuously differentiable and decreasing. Of course, g will depend on P. Let $h = \log \frac{1}{1-g^2}$, which is also decreasing. Note that

$$2\log T = \sum_p h\Bigl(\frac{p}{P}\Bigr) \geq \sum_{p\leq aP} h\Bigl(\frac{p}{P}\Bigr) \geq h(a)\pi(aP) \geq cah(a)\log T,$$

for P sufficiently large, for some constant $c > 0$. Thus $h(a) \leq C_a$ (independently of T).

Now, for $F\colon (0,\infty) \to [0,\infty)$ decreasing, it follows from the Prime Number Theorem in the form $\pi(x) = \mathrm{li}(x) + O(\frac{x}{(\log x)^2})$ that

$$\sum_{ax<p\leq bx} F\Bigl(\frac{p}{x}\Bigr) = \frac{x}{\log x}\int_a^b F + O\Bigl(\frac{xF(a)}{(\log x)^2}\Bigr), \tag{3.29}$$

where the implied constant is independent of F (and x). Thus

$$2\log T \geq \sum_{aP<p\leq AP} h\Bigl(\frac{p}{P}\Bigr) \sim \frac{P}{\log P}\int_a^A h \sim \log T \int_a^A h.$$

Since a and A are arbitrary, $\int_0^\infty h$ must exist and is at most 2. Also, by (3.29),

$$\sum_{aP<p\leq AP} \frac{\gamma_p}{p^\alpha} = \frac{1}{P^\alpha}\sum_{aP<p\leq AP} g\Bigl(\frac{p}{P}\Bigr)\Bigl(\frac{p}{P}\Bigr)^{-\alpha} \sim \frac{P^{1-\alpha}}{\log P}\int_a^A \frac{g(u)}{u^\alpha}\,du.$$

As a, A are arbitrary, it follows from above and (3.26), (3.27), (3.28) that

$$\log\frac{\|\varphi_\alpha a\|}{\|a\|} \leq \Bigl(\int_0^\infty \frac{g(u)}{u^\alpha}\,du + o(1)\Bigr)\frac{(\log T)^{1-\alpha}}{(\log\log T)^\alpha}.$$

Thus we need to maximise $\int_0^\infty g(u)u^{-\alpha}du$ subject to $\int_0^\infty h \le 2$ over all decreasing $g\colon (0,\infty) \to (0,1)$. Since h is decreasing, one finds that $xh(x) \to 0$ as $x \to \infty$ and as $x \to 0^+$.

For the supremum, we can consider only those g (and h) which are continuously differentiable and strictly decreasing, since we can approximate arbitrarily closely by such functions. On writing $g = s \circ h$, where $s(x) = \sqrt{1-e^{-x}}$, we have

$$\begin{aligned}\int_0^\infty \frac{g(u)}{u^\alpha}\,du &= \left[\frac{g(u)u^{1-\alpha}}{1-\alpha}\right]_0^\infty - \frac{1}{1-\alpha}\int_0^\infty g'(u)u^{1-\alpha}\,du \\ &= -\frac{1}{1-\alpha}\int_0^\infty s'(h(u))h'(u)u^{1-\alpha}\,du \\ &= \frac{1}{1-\alpha}\int_0^{h(0^+)} s'(x)l(x)^{1-\alpha}\,dx,\end{aligned}$$

where $l = h^{-1}$, since $\sqrt{u}g(u) \to 0$ as $u \to \infty$. The final integral is, by Hölder's inequality, at most

$$\left(\int_0^{h(0^+)} s'^{1/\alpha}\right)^\alpha \left(\int_0^{h(0^+)} l\right)^{1-\alpha}. \tag{3.30}$$

But $\int_0^{h(0^+)} l = -\int_0^\infty uh'(u)du = \int_0^\infty h \le 2$, so

$$\int_0^\infty \frac{g(u)}{u^\alpha}\,du \le \frac{2^{1-\alpha}}{1-\alpha}\left(\int_0^\infty s'^{1/\alpha}\right)^\alpha.$$

A direct calculation shows that[6] $\int_0^\infty (s')^{1/\alpha} = 2^{-1/\alpha}B(\frac{1}{\alpha}, 1-\frac{1}{2\alpha})$. This gives the upper bound.

The proof of the upper bound leads to the optimal choice for g and the lower bound. We note that we have equality in (3.30) if $l/(s')^{1/\alpha}$ is constant, i.e., $l(x) = cs'(x)^{1/\alpha}$ for some constant $c > 0$ — chosen so that $\int_0^\infty l = 2$. This means that we take

$$h(x) = (s')^{-1}\left(\left(\frac{x}{c}\right)^\alpha\right) = \log\left(\frac{1}{2} + \frac{1}{2}\sqrt{1+\left(\frac{c}{x}\right)^{2\alpha}}\right),$$

from which we can calculate g. In fact, the required lower bound can be found by taking a_n completely multiplicative, with a_p for p prime defined by

$$a_p = g_0\left(\frac{p}{P}\right),$$

where $P = \log T \log\log T$ and g_0 is the function

$$g_0(x) = \sqrt{1 - \frac{2}{1+\sqrt{1+(\frac{c}{x})^{2\alpha}}}},$$

[6] The integral is $2^{-1/\alpha}\int_0^\infty e^{-x/\alpha}(1-e^{-x})^{-1/2\alpha}dx = 2^{-1/\alpha}\int_0^1 t^{1/\alpha-1}(1-t)^{-1/2\alpha}dt$.

with $c = 2^{1+1/\alpha}/B(\frac{1}{\alpha}, 1 - \frac{1}{2\alpha})$. As such, by the same methods as before, we have $\|a\| = T^{1+o(1)}$ and

$$\log \frac{\|\varphi_\alpha a\|}{\|a\|} = \sum_p p^{-\alpha} g_0\Big(\frac{p}{P}\Big) + O(1) \sim \frac{P^{1-\alpha}}{\log P} \int_0^\infty \frac{g_0(u)}{u^\alpha}\, du.$$

By the choice of g_0, the integral on the right is $\dfrac{B(\frac{1}{\alpha}, 1 - \frac{1}{2\alpha})^\alpha}{(1-\alpha)2^\alpha}$, as required. □

Remark. These asymptotic formulae bear a strong resemblance to the (conjectured) maximal order of $|\zeta(\alpha + iT)|$. It is interesting to note that the bounds found here are just larger than what is known about the lower bounds for $Z_\alpha(T)$ (see the interlude on upper and lower bounds on $\zeta(s)$, especially items (b) and (c)). We note that the constant appearing in (3.24) is not Lamzouri's $C(\alpha)$ since, for α near $\frac{1}{2}$, the former is roughly $\frac{1}{\sqrt{\alpha - \frac{1}{2}}}$, while $C(\alpha) \sim \frac{1}{\sqrt{2\alpha - 1}}$.

Lower bounds for φ_α and some further speculations. We can study lower bounds of φ_α via the function

$$m_\alpha(T) = \inf_{\substack{a \in \mathcal{M}^2 \\ \|a\| = T}} \frac{\|\varphi_\alpha a\|}{\|a\|}.$$

Using very similar techniques, one obtains results analogous to Theorem 4.7:

$$\frac{1}{m_1(T)} = \frac{6e^\gamma}{\pi^2}(\log\log T + \log\log\log T + 2\log 2 - 1 + o(1))$$

and

$$\log \frac{1}{m_\alpha(T)} \sim \frac{B(\frac{1}{\alpha}, 1 - \frac{1}{2\alpha})^\alpha (\log T)^{1-\alpha}}{(1-\alpha)2^\alpha (\log\log T)^\alpha} \qquad \text{for } \tfrac{1}{2} < \alpha < 1.$$

We see that $m_\alpha(T)$ corresponds closely to the conjectured minimal order of $|\zeta(\alpha + iT)|$ (see [13] and [27]).

The above formulae suggest that the supremum (respectively infimum) of $\|\varphi_\alpha a\|/\|a\|$ with $a \in \mathcal{M}^2$ and $\|a\| = T$ are close to the supremum (resp. infimum) of $|\zeta_\alpha|$ on $[1, T]$. One could therefore speculate further that there is a close connection between $\|\varphi_\alpha a\|/\|a\|$ (for such a) and $|\zeta(\alpha + iT)|$.

Heuristically, we could argue as follows. Consider

$$\frac{1}{T}\int_0^T |\zeta(\alpha - it)|^2 \left|\sum_{n=1}^\infty a_n n^{it}\right|^2 dt. \tag{3.31}$$

This is less than

$$Z_\alpha(T)^2 \cdot \frac{1}{T}\int_0^T \left|\sum_{n=1}^\infty a_n n^{it}\right|^2 dt \sim Z_\alpha(T)^2 \|a\|^2,$$

by the Montgomery–Vaughan mean value theorem (under appropriate conditions). On the other hand, (3.31) is expected to be approximately

$$\frac{1}{T}\int_0^T \left|\sum_{n=1}^{\infty} b_n n^{it}\right|^2 dt \approx \sum_{n=1}^{\infty} |b_n|^2 = \|\varphi_\alpha a\|^2.$$

Putting these together gives

$$\frac{\|\varphi_\alpha a\|}{\|a\|} \lesssim Z_\alpha(T).$$

The left-hand side, as a function of $\|a\|$, can be made as large as $F(\|a\|)$, where $F(x) = \exp\{c_\alpha \frac{(\log x)^{1-\alpha}}{(\log\log x)^\alpha}\}$. If the above continues to hold for $\|a\|$ as large as T, then $M_\alpha(T) \le Z_\alpha(T)$ would follow. Even if it holds for $\|a\|$ as large as a smaller power of T, one would recover Montgomery's Ω-result.

Alternatively, considering (3.31) over $[T, 2T]$,

$$\frac{1}{T}\int_T^{2T} |\zeta(\alpha - it)|^2 \left|\sum_{n=1}^{\infty} a_n n^{it}\right|^2 dt = |\zeta(\alpha + it_0)|^2 \cdot \frac{1}{T}\int_T^{2T} \left|\sum_{n=1}^{\infty} a_n n^{it}\right|^2 dt \quad (3.32)$$

for some $t_0 \in [T, 2T]$ and, using the Montgomery–Vaughan mean value theorem (assuming it applies), this is approximately

$$|\zeta(\alpha + it_0)|^2 \sum_{n=1}^{\infty} |a_n|^2 = |\zeta(\alpha + it_0)|^2 \|a\|^2.$$

On the other hand, formally multiplying out the integrand, by writing $\zeta(\alpha - it) = \sum_{n=1}^{\infty} \frac{n^{it}}{n^\alpha}$ and formally multiplying out the integrand, the left-hand side of (3.32) becomes (heuristically)

$$\frac{1}{T}\int_T^{2T} \left|\sum_{n=1}^{\infty} b_n n^{it}\right|^2 dt \approx \sum_{n=1}^{\infty} |b_n|^2 = \|\varphi_\alpha a\|^2.$$

Equating these gives

$$\frac{\|\varphi_\alpha a\|}{\|a\|} \approx |\zeta(\alpha + it_0)|.$$

Clearly there are a number of problems with this. For a start, we need $\varphi_\alpha a \in l^2$. More importantly, the error term in the Montgomery–Vaughan theorem contains $\sum_{n=1}^{\infty} n|a_n|^2$, which may diverge. Also, a_n and hence $\|a\|$ may depend on T, and finally, the series for $\zeta(\alpha - it)$ doesn't converge for $\alpha \le 1$.

If $a_n = 0$ for $n > N$, the above argument can be made to work, even for N a (small) power of T (see for example [17]), but difficulties arise for larger powers of T.

There seem to be some reasons to believe that the error from the Montgomery–Vaughan theorem should be much smaller when considering products. These occur when a_n is multiplicative. For example (with $Q = \prod_{p \le P} p$ so that $\log Q = \theta(P) \sim P$ by the Prime Number Theorem),

$$\frac{1}{2T}\int_{-T}^{T}\left|\prod_{p\le P}(1+p^{it})\right|^2 dt = \frac{1}{2T}\int_{-T}^{T}\left|\sum_{d|Q} d^{it}\right|^2 dt = \sum_{d|Q} 1 + O\Big(\frac{1}{T}\sum_{d|Q} d\Big).$$

The 'main term' is $d(Q) = 2^{\pi(P)}$, while the error is at least $\frac{Q}{T}$. However the left-hand side is trivially at most $4^{\pi(P)}$, so the error dominates the other terms if $P > (1+\delta)\log T$. If, say, P is of order $\log T \log\log T$ (which is the range of interest), then $\pi(P) \asymp \log T$, so $2^{\pi(P)}$ is like a power of T, but Q is roughly like $T^{\log\log T}$ – far too large.

Thus it may be that for a_n completely multiplicative, it holds that

$$\frac{1}{T}\int_{T}^{2T} |\zeta(\alpha - it)|^2 \left|\prod_{p\le P}\frac{1}{1-a_p p^{it}}\right|^2 dt \sim |\zeta(\alpha + it_0)| \prod_{p \le P}\frac{1}{1-|a_p|^2}$$

for P up to $c \log T \log\log T$. This suggests the following might be true:

(a) *given $a \in \mathcal{M}^2$ with $\|a\| = T$, there exists $t \in [T, 2T]$ such that*

$$\frac{\|\varphi_\alpha a\|}{\|a\|} \approx |\zeta(\alpha + it)|.$$

(b) *given $T \ge 1$, there exists $a \in \mathcal{M}^2$ with $\|a\| = T$ such that*

$$\frac{\|\varphi_\alpha a\|}{\|a\|} \approx |\zeta(\alpha + iT)|.$$

Here, $\approx$ means something like log-asymptotic, $\sim$, or possibly even $=$. Thus (a) implies $M_\alpha(T) \lesssim Z_\alpha(T)$, while (b) implies the opposite. Together they would imply we can *encode* real numbers into $\mathcal{M}^2$-functions with equal l^2-norm, such that φ_α has a similar action as ζ_α.

Closure of $\mathcal{M}B^2_{\mathbb{N}}$? We finish these speculations with a final plausible conjecture regarding the closure under multiplication of functions in $B^2_{\mathbb{N}}$ with multiplicative coefficients.

Let $\mathcal{M}_0 B^2_{\mathbb{N}}$ denote the subset $\mathcal{M}B^2_{\mathbb{N}}$ of functions f for which $\hat{f} \in \mathcal{M}_0$. Recall that $\mathcal{M}^2_0$ is the subset of $\mathcal{M}^2$ for which $g \in \mathcal{M}^2 \Rightarrow f * g \in \mathcal{M}^2$. This suggests the following conjecture:

Conjecture. Let $f \in \mathcal{M}B^2_{\mathbb{N}}$ and $g \in \mathcal{M}_0 B^2_{\mathbb{N}}$. Then $fg \in \mathcal{M}B^2_{\mathbb{N}}$.

In particular, $\mathcal{M}B^2_{\mathbb{N}}\mathcal{M}_c B^2_{\mathbb{N}} = \mathcal{M}B^2_{\mathbb{N}}$. Since $\zeta_\alpha \in \mathcal{M}_c B^2_{\mathbb{N}}$ for $\alpha > \frac{1}{2}$, this would imply $\zeta_\alpha^k \in \mathcal{M}B^2_{\mathbb{N}}$ for every $k \in \mathbb{N}$ and $\alpha > \frac{1}{2}$, which implies the Lindelöf hypothesis.

5 Connections to matrices of the form $(f(ij/(i,j)^2))_{i,j\le N}$

The asymptotic formulae for $\Phi_\alpha(N)$ in Theorem 4.1 can be used to obtain information on the largest eigenvalue of certain arithmetical matrices. Various authors have discussed asymptotic estimates of eigenvalues and determinants of arithmetical matrices (see for example [6], [7], [26] to name just a few).

Let $A_N(f)$ denote the $N\times N$ matrix with ij^{th}-entry $f(i/j)$ if $j|i$ and zero otherwise. As noted in the introduction, these matrices behave much like Dirichlet series with coefficients $f(n)$; namely,

$$A_N(f)A_N(g) = A_N(f*g).$$

In particular, $A_N(f)$ is invertible if f has a Dirichlet inverse, i.e., $f(1)\neq 0$, in which case $A_N(f)^{-1} = A_N(f^{-1})$.

Suppose for simplicity that f is a real arithmetical function. (For complex values we can easily adjust.) Let

$$\Phi_f(N) = \sup_{\|a\|_2=1}\left(\sum_{n=1}^{N}|b_n|^2\right)^{\frac12}$$

where $b_n = \sum_{d|n} f(d)a_{n/d}$. Observe that $\Phi_f(N)^2$ is the largest eigenvalue of the matrix

$$A_N(f)^T A_N(f).$$

Indeed, we have

$$b_n^2 = \sum_{i,j|n} f\Big(\frac{n}{i}\Big)f\Big(\frac{n}{j}\Big)a_ia_j,$$

so that, on noting $i,j|n$ if and only if $[i,j]|n$ (where $[i,j]$ denotes the lcm of i and j)

$$\sum_{n=1}^{N} b_n^2 = \sum_{n=1}^{N}\sum_{[i,j]|n} f\Big(\frac{n}{i}\Big)f\Big(\frac{n}{j}\Big)a_ia_j = \sum_{i,j\le N} b_{ij}^{(N)}a_ia_j,$$

where (using $(i,j)[i,j]=ij$)

$$b_{ij}^{(N)} = \sum_{k\le\frac{N}{[i,j]}} f\Big(\frac{ki}{(i,j)}\Big)f\Big(\frac{kj}{(i,j)}\Big).$$

But $b_{ij}^{(N)}$ is also the ij^{th}-entry of $A_N(f)^TA_N(f)$, as an easy calculation shows. Thus

$$\Phi_f(N)^2 = \sup_{a_1^2+\dots+a_N^2=1}\sum_{i,j\le N} b_{ij}^{(N)}a_ia_j \tag{3.33}$$

is the largest eigenvalue of $A_N(f)^T A_N(f)$, i.e., $\Phi_f(N)$ the largest *singular value*[7] of $A_N(f)$. Thus an equivalent formulation of Corollary 3.2 for f supported on $\mathbb{N}$ is: *For $f \in l^1$ non-negative, the largest singular value of $A_N(f)$ tends to $\|f\|_1$.*

Now if f is completely multiplicative, then

$$b_{ij}^{(N)} = f\Big(\frac{ij}{(i,j)^2}\Big) \sum_{k \le \frac{N}{[i,j]}} f(k)^2,$$

which for large N is roughly $\|f\|_2^2 f(\frac{ij}{(i,j)^2})$ for $f \in l^2$. This suggests that the matrix $\big(f(\frac{ij}{(i,j)^2})\big)_{i,j \le N}$ has its largest eigenvalue close to $\Phi_f(N)^2/\|f\|_2^2$. This is indeed the case.

Corollary 5.1. *Let $f \in l^2$ be non-negative and completely multiplicative. Let Λ_N denote the largest eigenvalue of $\big(f(\frac{ij}{(i,j)^2})\big)_{i,j \le N}$. Then*

$$\frac{\Phi_f(N)^2}{\|f\|_2^2} \le \Lambda_N \le \frac{\Phi_f(N^3)^2}{\sum_{k=1}^N f(k)^2}.$$

In particular, for $f \in l^1$,

$$\lim_{N\to\infty} \Lambda_N = \frac{\|f\|_1^2}{\|f\|_2^2}.$$

Proof. We have

$$\Lambda_N = \sup_{a_1^2+\cdots+a_N^2=1} \sum_{i,j \le N} f\Big(\frac{ij}{(i,j)^2}\Big) a_i \overline{a_j} \tag{3.34}$$

When $f \ge 0$, the supremums in (3.33) and (3.34) are reached for $a_n \ge 0$. Thus,

$$\Phi_f(N)^2 \le \|f\|_2^2 \Lambda_N$$

follows immediately.

On the other hand, for $i, j \le N$, $[i,j] \le N^2$ so

$$\Phi_f(N^3)^2 \ge \sum_{i,j \le N} f\Big(\frac{ij}{(i,j)^2}\Big) a_i a_j \sum_{k \le N} f(k)^2.$$

Taking the supremum over all such a_n gives, $\Phi_f(N^3)^2 \ge \Lambda_N \sum_{k \le N} f(k)^2$, as required.

Finally, if $f \in l^1$, then $\Phi_f(N) \to \|f\|_1$ and so $\Lambda_N \to \frac{\|f\|_1^2}{\|f\|_2^2}$ follows. □

[7] The singular values of a matrix A are the square roots of the eigenvalues of $A^T A$ (or $A^* A$ if A has complex entries).

The approximate formulae for $\Phi_\alpha(N)$ in Theorem 4.1 lead to:

Corollary 5.2. *Let* $f(n) = n^{-\alpha}$ *and let* $\Lambda_N(\alpha)$ *denote the largest eigenvalue of* $\big(f(\frac{ij}{(i,j)^2})\big)_{i,j\le N}$. *Then*

$$\Lambda_N(1) = \frac{6}{\pi^2}(e^\gamma \log\log N + O(1))^2,$$
$$\log \Lambda_N(\alpha) \asymp \frac{(\log N)^{1-\alpha}}{\log\log N} \qquad \text{for } \tfrac{1}{2} < \alpha < 1,$$
$$\log \Lambda_N\Big(\frac{1}{2}\Big) \asymp \sqrt{\frac{\log N}{\log\log N}}.$$

Bibliography

[1] C. Aistleitner, Lower bounds for the maximum of the Riemann zeta function along vertical lines, arXiv:1409.6035.

[2] R. Balasubramanian and K. Ramachandra, On the frequency of Titchmarsh's phenomenom for $\zeta(s)$. III, *Proc. Indian Acad. Sci.* 86 A (1977), 341–351.

[3] A. S. Besicovitch, *Almost Periodic Functions*. Dover Publications, 1954.

[4] A. Bondarenko and K. Seip, Large GCD sums and extreme values of the Riemann zeta function, arXiv:1507.05840v1.

[5] A. Böttcher and B. Silbermann, *Introduction to Large Truncated Toeplitz Matrices*. Springer-Verlag, 1999.

[6] K. Bourque and S. Ligh, Matrices associated with classes of arithmetical functions. *J. Number Theory* 45 (1993), 367–376.

[7] D. A. Cardon, Matrices related to Dirichlet series. *J. Number Theory* 130 (2010), 27–39.

[8] D. W. Farmer, S. M. Gonek and C. P. Hughes, The maximum size of L-functions. *J. Reine Angew. Math.* 609 (2007), 215–236.

[9] I. Gohberg, On an application of the theory of normed rings to singular integral equations. *Usp. Mat. Nauk* 7 (1952), 149–156.

[10] I. Gohberg, S. Goldberg and M. A. Kaashoek, *Classes of Linear Operators Vol. I*. Birkhäuser, 1990.

[11] S. M. Gonek, Finite Euler products and the Riemann Hypothesis, *Trans. Amer. Math. Soc.* 364 (2012), no. 4, 2157–2191.

[12] S. M. Gonek and J. P. Keating, Mean values of finite Euler products. *J. London Math. Soc.* 82 (2010), 763–786.

[13] A. Granville and K. Soundararajan, Extreme values of $|\zeta(1+it)|$. In *Ramanujan Math. Soc. Lect. Notes Ser 2*, Ramanujan Math. Soc., Mysore, 2006, 65–80.

[14] P. Hartman, On completely continuous Hankel operators. *Proc. Am. Math. Soc.* 9 (1958), 862–866.

[15] E. Hewitt and J. H. Williamson, Note on absolutely convergent Dirichlet series. *Proc. Am. Math. Soc.* 8 (1957), 863–868.

[16] T. W. Hilberdink, Determinants of multiplicative Toeplitz matrices. *Acta Arith.* 125 (2006), 265–284.

[17] T. W. Hilberdink, An arithmetical mapping and applications to Ω-results for the Riemann zeta function. *Acta Arith.* 139 (2009), 341–367.

[18] T. W. Hilberdink, 'Quasi'-norm of an arithmetical convolution operator and the order of the Riemann zeta function. *Funct. Approx.* 49 (2013), 201–220.

[19] M. Huxley, Exponential sums and the Riemann zeta function V. *PLMS* 90 (2005), 1–41.

[20] T. J. Jarvis, A dominant negative eigenvalue of a matrix of Redheffer. *Linear Algebra Appl.* 142 (1990), 141–152.

[21] B. Jessen and H. Tornehave, Mean motions and zeros of almost periodic functions. *Acta Math.* 77 (1945), 137–279.

[22] J. Keating and N. Snaith, *Random matrix theory and some zeta-function moments*. Lecture at the Erwin Schrödinger Institute (1998).

[23] M. Krein, Integral equations on a half-line with kernel depending upon the difference of the arguments. *Am. Math. Soc. Transl.* (2) 22 (1962), 163–288.

[24] Y. Lamzouri, On the distribution of extreme values of zeta and L-functions in the strip $\frac{1}{2} < \sigma < 1$. *Int. Math. Res. Nat.*, to appear; arXiv:1005.4640v2 [math.NT].

[25] N. Levinson, Ω-theorems for the Riemann zeta function. *Acta Arith.* 20 (1972), 319–332.

[26] P. Lindqvist and K. Seip, Note on some greatest common divisor matrices. *Acta Arith.* 84 (1998), 149–154.

[27] H. L. Montgomery, Extreme values of the Riemann zeta-function. *Comment. Math. Helv.* 52 (1977), 511–518.

[28] Z. Nehari, On bounded bilinear forms. *Ann. Math.* 65 (1957), 153–162.

[29] R. M. Redheffer, Eine explizit lösbare Optimierungsaufgabe. *Int. Schriftenr. Numer. Math.* 36 (1977).

[30] L. Rodman, I. M. Spitkowsky and H. J. Woerdeman, Fredholmness and Invertibility of Toeplitz operators with matrix almost periodic symbols. *Proc. Am. Math. Soc.* 130 (2001), 1365–1370.

[31] K. Soundararajan, Extreme values of zeta and L-functions. *Math. Ann.* 342 (2008), 467–486.

[32] A. E. Taylor, *Introduction to Functional Analysis*. Wiley and Sons, 1958.

[33] E. C. Titchmarsh, *The Theory of the Riemann Zeta-function*. Second edition, Oxford University Press, 1986.

[34] O. Toeplitz, Zur Theorie der quadratischen und bilinearen Formen von unendlichvielen Veränderlichen. *Math. Ann.* 70 (1911), 351–376.

[35] O. Toeplitz, Zur Theorie der Dirichletschen Reihen. *Am. J. Math.* 60 (1938), 880–888.

Chapter 4

Arithmetical topics in algebraic graph theory

Jürgen Sander

Contents

1 Introduction

In comparison with geometry, number theory, analysis, or even algebra and topology, **graph theory** is a rather young mathematical discipline. It has undergone crucial developments with important theoretical advances as well as closer connections to other mathematical fields and applications in extra-mathematical areas only within the last 80 years. The term **graph** was introduced by Sylvester in 1878 in an article published in “Nature”, the by far most influential scientific magazine worldwide.

By now graph theory is without any doubt an important mathematical discipline in its own right with prestigious topics, results and open questions, as for instance the four color problem, just to name the most famous one. At the same time, it is influential for many other mathematical fields, not only in discrete mathematics, but also for areas such as algebra, topology and even probability theory. Naturally, the development of graph theory benefited from the rapid progress in computers and computer science in the 20th century.

However, graph theory has yet another facet to it, namely it may serve as a mathematical “playground” in the following sense. Quite a few mathematical concepts in other fields, as for instance topological items or the variety of zeta functions, related to number theory, algebra, differential geometry, or other subjects, have a counterpart in the world of (finite) graphs. Looking at and experimenting with these somewhat discrete relatives sometimes helps to better understand the original objects.

Algebraic graph theory – nomen est omen – connects algebra and graph theory. *Cayley graphs*, encoding structural features of groups in correlated graphs, are a prominent example studied in this rather young mathematical field. Besides the necessary basics of algebraic graph theory, these lecture notes will present three topics connecting graph theory and arithmetic.

2 Some (algebraic) graph theory

Here we give a short introduction to the concepts of graph theory that will be used in the sequel. Besides some standard basics like cycle graphs, trees, regular graphs, spanning subgraphs, etc., which can be found in almost any introductory book on graph theory (e.g. [12]) and are mainly included to set up the notation, the reader will find more specialized results on circulant graphs and Cayley graphs. The introduction to algebraic graph theory deals with the adjacency matrix and the spectrum of graphs, in particular the spectrum of circulant and integral circulant graphs.

2.1 Basics and some specialities on graphs

Definition 2.1.

- A (*finite*) *graph* $\mathcal{G} = (V, E)$ consists of a (finite) non-empty set V of *vertices* and a set $E \subseteq V \times V$ of *edges*. The *order* of $\mathcal{G}$ is the number of vertices $|V|$ in $\mathcal{G}$.

  We usually write $v_1 v_2$ rather than (v_1, v_2), and for $v_1 v_2 \in E$ we say that v_1 and v_2 are *adjacent* or *neighbors* in $\mathcal{G}$, written $v_1 \sim v_2$.
- The graph $\mathcal{G}$ is called *undirected* if E is symmetric, i.e., if $v_1 \sim v_2 \;\Leftrightarrow\; v_2 \sim v_1$ for all $v_1, v_2 \in V$, and otherwise *directed*. For undirected graphs we shall not distinguish between $v_1 v_2$ and $v_2 v_1$ (here $v_1 v_2$ represents a set rather than an ordered pair).
- Edges of type $vv \in E$ are called *loops*. A *loopfree* graph has no loops.

For several applications it is useful to allow graphs with *multiple edges* between vertices, so-called *multigraphs*. We call graphs without loops or multiple edges *simple*.

Assumption. Unless explicitly stated otherwise, the graphs $\mathcal{G} = (V, E)$ we consider are always **simple**, **finite**, and **undirected**.

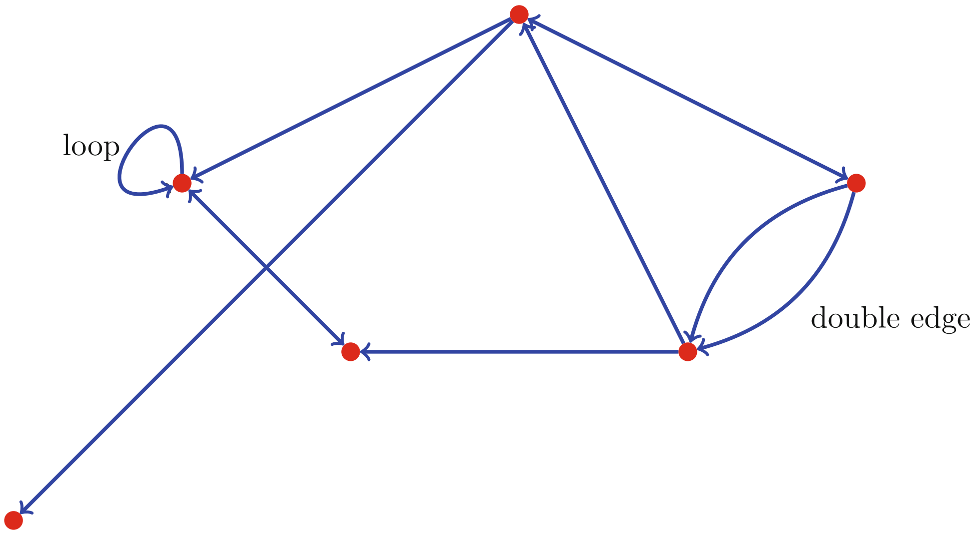


Figure 4.1. Directed multigraph with loop and double edge

Example 2.2. Let n be a positive integer.

(i) The *complete graph* $\mathcal{K}_n = (V, E)$ *of order* n has $|V| = n$ and $E = V \times V \setminus \{vv : v \in V\}$. Clearly, $|E| = \frac{n(n-1)}{2}$.

(ii) The *cycle graph* $\mathcal{C}_n = (V, E)$ of order n has $V = \{v_1, v_2, \ldots, v_n\}$, say, and $E = \{v_1v_2, v_2v_3, \ldots, v_{n-1}v_n, v_nv_1\}$.

(iii) The graph $(\mathbb{Z}_{10}, E_{\{1,3,7,9\}})$ with $\mathbb{Z}_{10} := \mathbb{Z}/10\mathbb{Z}$ (cf. Fig. 4.2) is defined by

$$E_{\{c_1,\ldots,c_r\}} := \{ab \in \mathbb{Z}_{10} \times \mathbb{Z}_{10} : a - b \equiv c_j \bmod 10 \text{ for some } j \in \{1, \ldots, r\}\}.$$

It is an example of a *circulant graph* (see Example 2.19), as well as of a *Cayley graph* (see Example 2.12 (iii)).

The structure of a graph does not depend on how the vertices are named. We already made use of this in the preceding examples by speaking of **the** complete graph and **the** cycle graph of order n, respectively, without specifying the elements of V. Accordingly, two graphs $\mathcal{G}_1 = (V_1, E_1)$ and $\mathcal{G}_2 = (V_2, E_2)$ are called *isomorphic*, written $\mathcal{G}_1 \simeq \mathcal{G}_2$, if there is a bijective map $\varphi : V_1 \to V_2$ such that $v_1 \sim v_2$ in $\mathcal{G}_1$ if and only if $\varphi(v_1) \sim \varphi(v_2)$ in $\mathcal{G}_2$. From now on, we shall not distinguish between isomorphic graphs.

Definition 2.3. Let $\mathcal{G} = (V, E)$ and $\mathcal{G}' = (V', E')$ be graphs with $V' \subseteq V$.

- $\mathcal{G}'$ is called a *subgraph* of $\mathcal{G}$ if $E' \subseteq (V' \times V') \cap E$.
- In the special case where $E' = (V' \times V') \cap E$, we say that $\mathcal{G}'$ is *induced by its vertex set* V'.
- A subgraph $\mathcal{G}'$ of $\mathcal{G}$ is called a *spanning subgraph* if $V' = V$ (cf. Fig. 4.3).

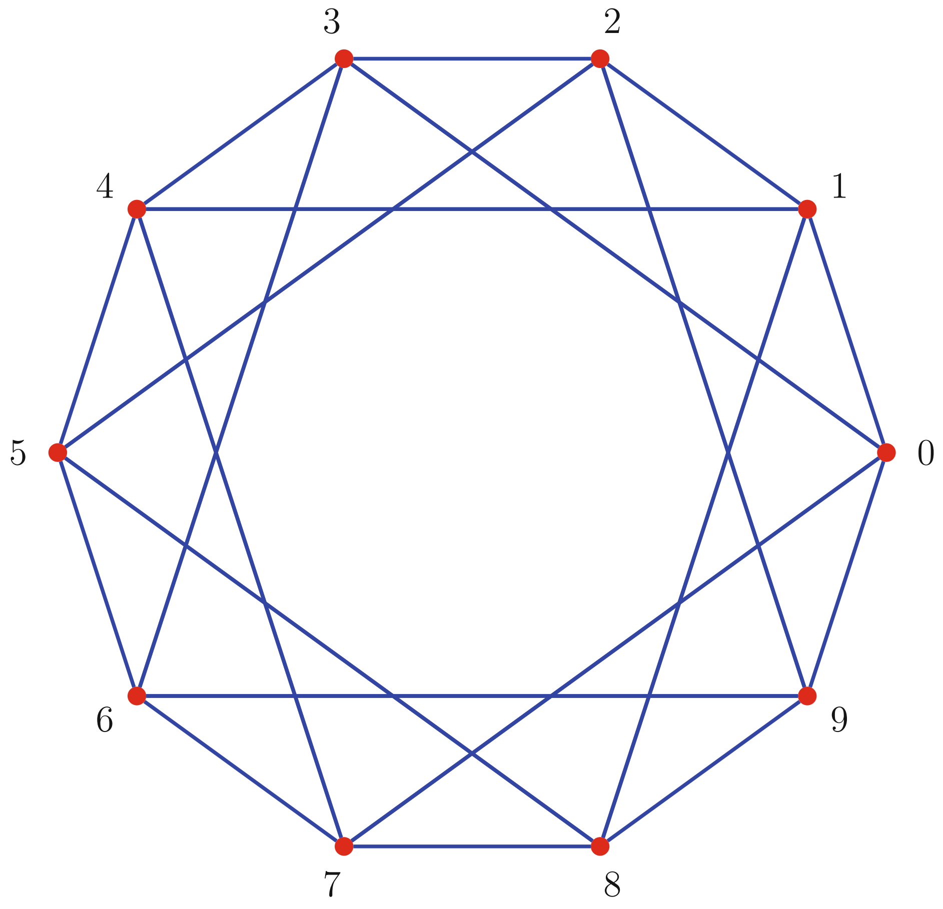


Figure 4.2. Circulant Cayley graph ($\mathbb{Z}_{10}$, $E_{\{1,3,7,9\}}$)

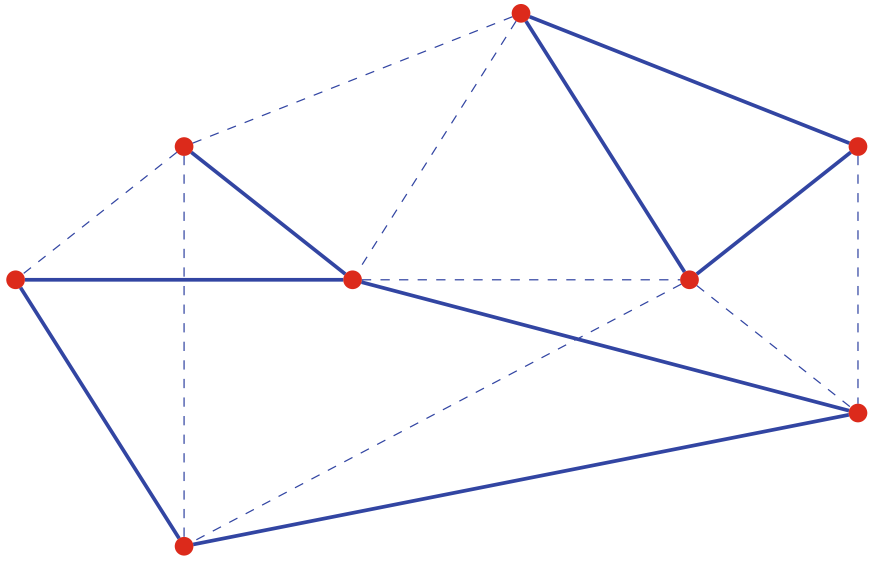


Figure 4.3. Graph with (disconnected) spanning subgraph (solid lines)

An important type of subgraph in a given graph is a *clique*, i.e., a subgraph which is a complete graph.

Definition 2.4. Let $\mathcal{G} = (V, E)$ be a (directed) graph.

- A (directed) *path* or *walk* from $u \in V$ to $v \in V$ of length k in $\mathcal{G}$ is a sequence of $k+1$ vertices $u = v_0, v_1, v_2, \ldots, v_{k-1}, v_k = v \in V$ such that $v_{i-1}v_i \in E$ for $i = 1, 2, \ldots, k$ (or, equivalently, the sequence of the corresponding k edges). The path is called *simple* if $v_i \neq v_j$ for $i = 1, 2, \ldots, k-1$ and all j. For a path $P = (v_0, v_1, v_2, \ldots, v_{k-1}, v_k)$, its length is denoted by $\nu(P) := k$.
- A (directed) *cycle* or *closed walk* in $\mathcal{G}$ is a (directed) path from v to v for some $v \in V$.
- If $C = (v_0, v_1, \ldots, v_{k-1}, v_0)$ is a cycle in $\mathcal{G}$, then

$$C^n := (\underbrace{v_0, v_1, \ldots, v_{k-1}, v_0, v_1, \ldots, v_{k-1}, \ \ldots \ , v_0, v_1, \ldots, v_{k-1}}_{n \text{ times}}, v_0)$$

  is called the n-th *power* of C.

Definition 2.5. Let $\mathcal{G} = (V, E)$ be a graph.

- $\mathcal{G}$ is called *connected* if there is a path from u to v for any vertices $u \neq v$ in $\mathcal{G}$; otherwise $\mathcal{G}$ is called *disconnected*.
- If $\mathcal{G}$ is disconnected, its induced connected subgraphs $\mathcal{G}_i = (V_i, E_i)$ with pairwise disjoint $V_i \subset V$ satisfying $\bigcup V_i = V$ are called the *components* of $\mathcal{G}$.

All graphs in Example 2.2 are connected and contain cycles.

Example 2.6. The graph $(\mathbb{Z}_{10}, E_{\{2,8\}})$ is disconnected with two components.

If a graph $\mathcal{G}$ contains a cycle, we call $\mathcal{G}$ *cyclic*, otherwise we call it *acyclic*. We introduce the very important subclass of graphs which are connected, but acyclic.

Definition 2.7. A *tree* is a connected acyclic graph (cf. Fig. 4.5).

Exercise 2.8. Show that the following assertions are equivalent for a graph $\mathcal{T} = (V, E)$:

(i) $\mathcal{T}$ is a tree.

(ii) There is a unique path in $\mathcal{T}$ between any two distinct vertices of $\mathcal{T}$.

(iii) $\mathcal{T}$ is *maximally acyclic*, i.e., $\mathcal{T}$ is acyclic, but $\mathcal{T}' := (V, E \cup \{e\})$ is cyclic for all $e \in (V \times V) \setminus E$.

(iv) $\mathcal{T}$ is *minimally connected*, i.e., $\mathcal{T}$ is connected, but $\mathcal{T}'' := (V, E \setminus \{e\})$ is disconnected for every $e \in E$. This means that $|E| = |V| - 1$.

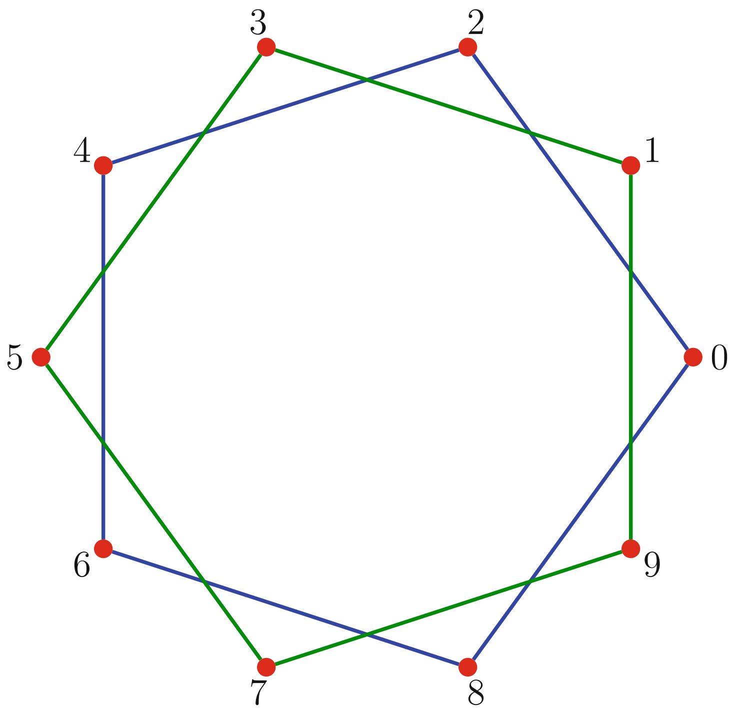


Figure 4.4. Disconnected Cayley graph $(\mathbb{Z}_{10}, E_{\{2,8\}})$ with two components

Proposition 2.9. *Every connected graph $\mathcal{G}$ contains a* spanning tree*, which by definition means that there is a spanning subgraph of $\mathcal{G}$ which is a tree.*

Proof. $\mathcal{G}$ contains a connected spanning subgraph, e.g., $\mathcal{G}$ itself. Now assume that $\mathcal{G}'$ is a connected spanning subgraph of $\mathcal{G}$ with a minimal number of edges. Then the equivalence of (i) and (iv) in Exercise 2.8 tells us that $\mathcal{G}'$ is a tree. □

Definition 2.10. Let $\mathcal{G} = (V, E)$ be a graph.

- The *degree* of $v \in V$ is $d(v) := |\{u \in V : u \sim v\}|$, i.e., the number of neighbors of v.
- A vertex of degree 1 is called a *leaf*.
- If all vertices in V have the same degree (let's say k), then $\mathcal{G}$ is called *regular* or, more precisely, k*-regular*.

Exercise 2.11. Every tree with more than two vertices has at least two leaves.

Example 2.12.

(i) Each cycle graph $\mathcal{C}_n$ is 2-regular.

(ii) The complete graph $\mathcal{K}_n$ is $(n-1)$-regular.

(iii) The circulant graph $(\mathbb{Z}_{10}, E)$ from Example 2.2 is 4-regular.

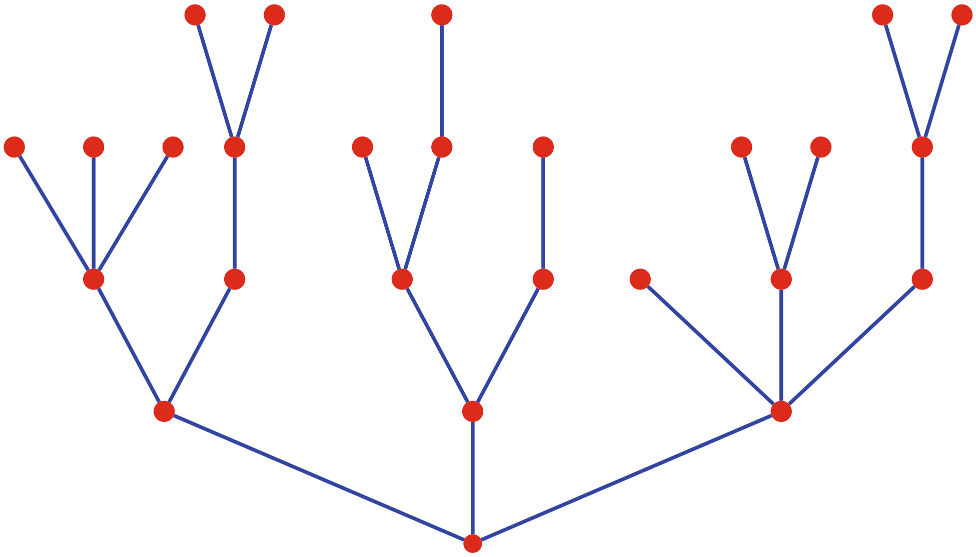


Figure 4.5. Tree

Definition 2.13. Let G be a finite group and $S \subseteq G$. We usually assume that S is a generating set of G and that S is *symmetric*, i.e., $s \in S$ implies $s^{-1} \in S$.

(i) Then the corresponding *Cayley graph* $\mathrm{Cay}(G, S) = (V, E)$ is defined by $V := G$ and $E := \{(g, gs) : \ g \in G, \ s \in S\}$.

(ii) A Cayley graph $\mathrm{Cay}(\mathbb{Z}_n, \mathbb{Z}_n^*)$ is called *unitary*, where $\mathbb{Z}_n := \mathbb{Z}/n\mathbb{Z}$ is the additive group of residues mod n for some positive integer n and $\mathbb{Z}_n^* = \{1 \leq a \leq n : \ (a, n) = 1\}$ is the multiplicative unit group in the residue class ring $\mathbb{Z}_n$.

Example 2.14. Clearly, $(\mathbb{Z}_{10}, E_{\{1,3,7,9\}}) = \mathrm{Cay}(\mathbb{Z}_{10}, \{\pm 1, \pm 3\})$ (cf. Example 2.2 (iii)) is a unitary Cayley graph, and $(\mathbb{Z}_{10}, E_{\{2,8\}}) = \mathrm{Cay}(\mathbb{Z}_{10}, \{\pm 2\})$ (cf. Example 2.6) is a disconnected Cayley graph.

Exercise 2.15. Show that

(i) $\mathrm{Cay}(G, S)$ is connected if and only if S is a generating set.

(ii) $\mathrm{Cay}(G, S)$ is undirected if and only if S is symmetric.

(iii) $\mathrm{Cay}(G, S)$ is regular of degree $|S|$.

Definition 2.16. A graph $\mathcal{G} = (V, E)$ is called *bipartite* if V is the union of two disjoint subsets V_1 and V_2 such that $E \subseteq (V_1 \times V_2) \cup (V_2 \times V_1)$ (cf. Fig. 4.6).

Exercise 2.17.

(i) Show that a bipartite graph cannot contain a cycle of odd length.

(ii) Prove that bipartite graphs are characterized by the property in (i).

Hint: Assume the graph is connected and consider a spanning tree.

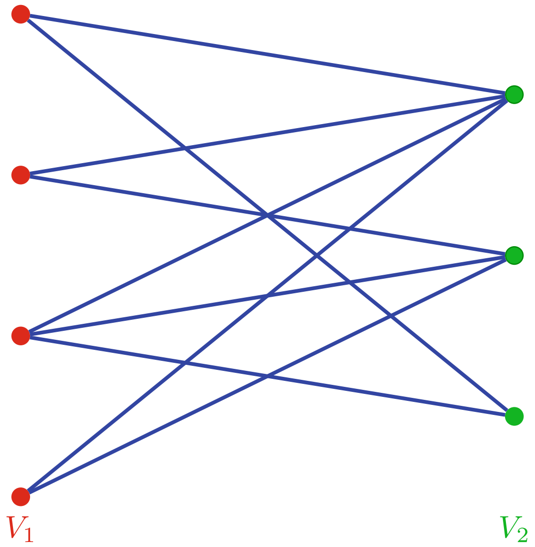


Figure 4.6. A bipartite graph

2.2 Some algebraic graph theory Much more than what we need from algebraic graph theory can be found in [7] of [14]. A tool which uniquely determines a graph and, more importantly, enables us to apply (linear) algebra to examine the graph and its properties is the adjacency matrix.

Definition 2.18. Let $\mathcal{G} = (V, E)$ be a graph with $V = \{v_1, \ldots, v_n\}$, which may be directed and is allowed to have loops. The *adjacency matrix* $A_{\mathcal{G}} = (a_{ij})_{n \times n}$ of $\mathcal{G}$ is defined by setting

$$a_{ij} = \begin{cases} 0, & \text{if } v_i \not\sim v_j\,, \\ 1, & \text{if } v_i \sim v_j \text{ for } i \neq j\,, \\ 2, & \text{if } v_i \sim v_i \text{ for } i = j\,. \end{cases}$$

Example 2.19. Consider the Cayley graph $\mathrm{Cay}(\mathbb{Z}_{10}, \{\pm 1, \pm 3\})$ with $\mathbb{Z}_{10} \simeq \{0, 1, 2, \ldots, 9\}$ from Example 2.2 (iii). Then

$$A_{\mathrm{Cay}(\mathbb{Z}_{10},\{\pm 1,\pm 3\})} = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

Exercise 2.20. Let $\mathcal{G}$ be an undirected graph, hence $A_{\mathcal{G}}$ is a real symmetric matrix.

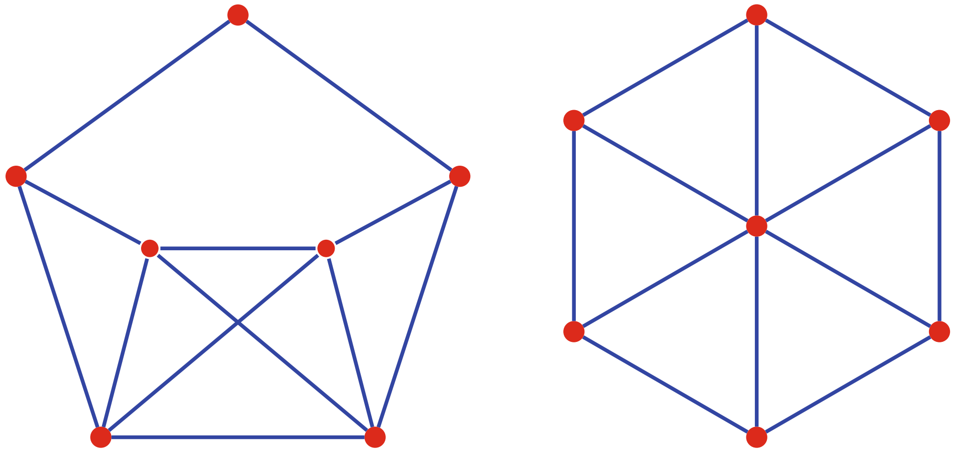


Figure 4.7. Two non-isomorphic cospectral graphs

(i) Let x and y be eigenvectors of $A_{\mathcal{G}}$ corresponding to different eigenvalues. Show that x and y are orthogonal.

(ii) Show that all eigenvalues of $A_{\mathcal{G}}$ are real.

(iii) Verify that all rational eigenvalues of $A_{\mathcal{G}}$ are integers.

Definition 2.21. Let $\mathcal{G} = (V, E)$ be a graph. The *eigenvalues of* $\mathcal{G}$ are the eigenvalues of $A_{\mathcal{G}}$. We call the multiset (= set with repetitions) of all eigenvalues of G the *spectrum* $\operatorname{Spec}(\mathcal{G}) := \operatorname{Spec}(A_{\mathcal{G}})$ of $\mathcal{G}$ (where $\operatorname{Spec}(\mathcal{G}) \subset \mathbb{R}$ by Exercise 2.20).

We have seen in Section 2.1 that renaming the vertices of a graph $\mathcal{G}_1$ gives an isomorphic graph $\mathcal{G}_2$. Usually $A_{\mathcal{G}_1} \neq A_{\mathcal{G}_2}$, but we have

Proposition 2.22. *If two graphs* $\mathcal{G}_1$ *and* $\mathcal{G}_2$ *are isomorphic, then* $\operatorname{Spec}(\mathcal{G}_1) = \operatorname{Spec}(\mathcal{G}_2)$.

Proof. Isomorphic graphs have the same number of vertices, thus assume that the sets V_1 and V_2 of vertices of $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively, are $V_1 = \{v_1, \dots, v_n\}$ and $V_2 = \{u_1, \dots, u_n\}$. Since $\mathcal{G}_1 \simeq \mathcal{G}_2$, we have $u_i = v_{\sigma(i)}$ for $i = 1, \dots, n$ with some permutation σ on $\{1, 2, \dots, n\}$. If P_σ denotes the $n \times n$ permutation matrix of σ, we have $P_\sigma^t A_{\mathcal{G}_1} P_\sigma = A_{\mathcal{G}_2}$. So $A_{\mathcal{G}_1}$ and $A_{\mathcal{G}_2}$ are similar matrices, and therefore they have the same characteristic polynomial, hence the same eigenvalues. □

Exercise 2.23.

(i) Show that $\operatorname{Spec}(\mathcal{G}_1) = \operatorname{Spec}(\mathcal{G}_2)$ (therefore $\mathcal{G}_1$ and $\mathcal{G}_2$ are called *cospectral*, or *isospectral*), but $\mathcal{G}_1 \not\simeq \mathcal{G}_2$ for the two graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ in Fig. 4.7. This means that the spectrum of a graph does not determine the structure of the graph. In other words, “you cannot hear the shape of a graph”, varying a famous question of Marc Kac posed in 1966.

In particular, the graph on the right obviously is *planar*, i.e., it can be embedded in the plane, while the graph on the left is not planar.

(ii) What was the famous question of Marc Kac referred to in (i)?

(iii) Look at some examples to find out that isomorphic graphs usually **do not** have the same eigenspaces!

A nice property of the adjacency matrix and its powers for a directed graph is given by the following result, which is proved by straightforward induction.

Proposition 2.24. *If $\mathcal{G} = (V, E)$ is a directed graph, then the number of (simple and nonsimple) directed paths of length ℓ from $u \in V$ to $v \in V$ is $(A_{\mathcal{G}}^{\ell})_{uv}$.*

Exercise 2.25. Let $\mathcal{G}$ be a graph (loopfree and undirected) with e edges and t triangles.

(i) Show that $\sum_{\lambda \in \mathrm{Spec}(\mathcal{G})} \lambda = \mathrm{tr} A_{\mathcal{G}} = 0$, where tr denotes the trace of a matrix.

The sum of the absolute values of the eigenvalues $\mathcal{E}(\mathcal{G}) := \sum_{\lambda \in \mathrm{Spec}(\mathcal{G})} |\lambda|$ is called the *energy* of $\mathcal{G}$.

(ii) Use Proposition 2.24 to prove that $\mathrm{tr} A_{\mathcal{G}}^2 = 2e$ and $\mathrm{tr} A_{\mathcal{G}}^3 = 6t$.

(iii) What about the corresponding problem for quadrilaterals?

The spectrum of a graph reflects quite a few of its properties. We are particularly interested in regular graphs.

Proposition 2.26. *Let $\mathcal{G}$ be a k-regular graph.*

(i) *Then $k \in \mathrm{Spec}(\mathcal{G})$, and $|\lambda| \le k$ for all $\lambda \in \mathrm{Spec}(\mathcal{G})$.*

(ii) *If $\mathcal{G}$ is connected, then k has multiplicity* 1 *in* $\mathrm{Spec}(\mathcal{G})$.

(iii) *If $\mathcal{G}$ is connected, then $-k \in \mathrm{Spec}(\mathcal{G})$ if and only if $\mathcal{G}$ is bipartite.*

Proof. Let $\mathcal{G} = (V, E)$ with $V = \{v_1, \ldots, v_n\}$. Since $\mathcal{G}$ is k-regular, every row of $A_{\mathcal{G}}$ contains k entries 1 and $n - k$ entries 0. Hence

$$A_{\mathcal{G}} \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} k \\ k \\ \vdots \\ k \end{pmatrix} = k \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix},$$

i.e., $k \in \mathrm{Spec}(\mathcal{G})$.

Let $\lambda \in \mathrm{Spec}(\mathcal{G})$ be arbitrary, hence $A_{\mathcal{G}} x = \lambda x$ for some $x = (x_1, \ldots, x_n) \neq 0$. Assume that $|x_m|$ is the largest among all $|x_i|$. Then

$$|\lambda x_m| = |(A_{\mathcal{G}} x)_m| = \Big| \sum_{v_i \sim v_m} x_i \Big| \le k |x_m|, \tag{4.1}$$

thus $|\lambda| \le k$, which proves (i).

To verify (ii), it remains to show that k has multiplicity 1 in $\mathrm{Spec}(\mathcal{G})$. Assume that $A_{\mathcal{G}} x = kx$ for some $x \neq 0$. Let again $|x_m|$ be the largest and w.l.o.g.

$x_m > 0$ (otherwise consider $-x$). It follows from (4.1) (without absolute values) that $\sum_{v_i \sim v_m} x_i = kx_m$. Since v_m has k neighbors, the sum has k terms and thus all corresponding summands x_i must be equal to x_m. Since $\mathcal{G}$ is connected, iteration of the argument finally yields $x_i = x_m$ for all i. This means that $(1, 1, \dots, 1)^t$ spans the eigenspace corresponding to $\lambda = k$.

We are left with (iii). First assume that $A_{\mathcal{G}}x = -kx$ for some $x \neq 0$. Again we may assume w.l.o.g. that some $x_m > 0$ has maximal $|x_m|$. Moreover, we may also assume that $x_i > 0$ for $1 \leq i \leq s$ and $x_i < 0$ for $t + 1 \leq i \leq n$, while $x_{s+1} = x_{s+2} = \dots = x_t = 0$. As in (4.1) we conclude that

$$-kx_m = \sum_{\substack{i=1 \\ v_i \sim v_m}}^{s} x_i + \sum_{\substack{i=t+1 \\ v_i \sim v_m}}^{n} x_i \geq \sum_{\substack{i=t+1 \\ v_i \sim v_m}}^{n} x_i \geq -kx_m.$$

This means we have in fact equality everywhere, hence $v_i \not\sim v_m$ for $1 \leq i \leq s$ and $x_i = -x_m$ for all $v_i \sim v_m$. Iterating this argument, each neighbor v_j, say, of a neighbor v_i of v_m has again $x_j = -x_i = x_m$. Since $\mathcal{G}$ is connected, V splits into two sets, $V_1 := \{v_i : x_i = x_m\}$ and $V_2 := \{v_i : x_i = -m\}$, such that $u \not\sim v$ for any $u, v \in V_1$ or any $u, v \in V_2$. This means that $\mathcal{G}$ is bipartite.

To prove the opposite direction of (iii) we observe that after reordering the vertices a connected bipartite graph $\mathcal{G}$ typically has an adjacency matrix of type

$$A_{\mathcal{G}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ * & * & * & * & * & * & 0 & 0 & 0 & 0 \\ * & * & * & * & * & * & 0 & 0 & 0 & 0 \\ * & * & * & * & * & * & 0 & 0 & 0 & 0 \\ * & * & * & * & * & * & 0 & 0 & 0 & 0 \end{pmatrix},$$

where the entries in the $*$ positions are 0 or 1. Now it is easy to see that

$$\begin{aligned} A_{\mathcal{G}} \cdot (1, \dots, 1, -1, \dots, -1)^t &= (-k, \dots, -k, k, \dots, k)^t \\ &= (-k)(1, \dots, 1, -1, \dots, -1)^t, \end{aligned}$$

i.e., $-k \in \operatorname{Spec}(\mathcal{G})$. □

Beyond the results of Proposition 2.26 the following facts are well known:

- A k-regular graph has as many components as the multiplicity of its eigenvalue k.
- $\operatorname{Spec}(\mathcal{G})$ is symmetric about the origin, i.e., the multiplicities of λ and $-\lambda$ in $\operatorname{Spec}(\mathcal{G})$ are the same, if and only if $\mathcal{G}$ is bipartite (this uses the Perron–Frobenius Theorem: see e.g. [14], Theorem 8.8.1 on the eigenvalues and eigenvectors of nonnegative matrices; also cf. Theorem 3.24).

Exercise 2.27. We know that $k \in \mathrm{Spec}(\mathcal{G})$ for a k-regular graph $\mathcal{G}$. Show that $\mathcal{G}$ necessarily is connected if k has multiplicity 1 in $\mathrm{Spec}(\mathcal{G})$.

Definition 2.28. A graph $\mathcal{G}$ is called *circulant* if $A_{\mathcal{G}}$ is a *circulant matrix* (briefly: a *circulant*, cf. [11]), i.e. a square matrix whose rows are obtained by cylic right shifts of the first row.

Example 2.29. Clearly, $\mathrm{Cay}(\mathbb{Z}_{10}, \{\pm 1, \pm 3\})$, the cycle graphs $\mathcal{C}_n$ and the complete graphs $\mathcal{K}_n$ are examples of circulant graphs.

Proposition 2.30 (Spectrum of $\mathcal{C}_n$). *Let n be a positive integer and $\omega_j := \exp(\frac{2\pi i j}{n})$ for $j = 0, 1, \ldots, n-1$. We have*

(i) $\mathrm{Spec}(\mathcal{C}_n) = \{\omega_j + \frac{1}{\omega_j} : \ j = 0, 1, \ldots, n-1\}$.

(ii) *The largest eigenvalue of $\mathcal{C}_n$ is* 2 (*with multiplicity* 1)*, and the second largest is* $2\cos\frac{(n-1)\pi}{n}$ (*with multiplicity* 2)*. The smallest eigenvalue is* -2 (*with multiplicity* 1) *for even n, and it is* $-2\cos\frac{(n-1)\pi}{n}$ (*with multiplicity* 2) *for odd n.*

(iii) *$\mathcal{C}_n$ is bipartite if and only if n is even.*

Proof. A helpful observation is that $A_{\mathcal{C}_n} = P + P^{-1}$, where P is the permutation matrix of the permutation defined by a cyclic left shift. If ω is any n-th root of unity, then it is obvious that the vector $(1, \omega, \omega^2, \ldots, \omega^{n-1})^t$ is an eigenvector of P with eigenvalue ω as well as an eigenvector of P^{-1} with eigenvalue $\frac{1}{\omega}$. Since there are exactly n n-th roots of unity, namely $\omega = \omega_j$ for $j = 0, 1, \ldots, n-1$, identity (i) follows.

By (i), the largest eigenvalue is $\omega_0 + \frac{1}{\omega_0} = 2$ (with multiplicity 1), and the second largest is $2\cos\frac{(n-1)\pi}{n} = \omega_1 + \frac{1}{\omega_1} = \omega_{n-1} + \frac{1}{\omega_{n-1}}$ (with multiplicity 2). The smallest eigenvalue is -2 (with multiplicity 1) for even n, and it is $2\cos\frac{(n-1)\pi}{n}$ (with multiplicity 2) for odd n.

(iii) follows from (ii) and Proposition 2.26 (iii). □

Arguing as in the preceding Proposition 2.30 one can show

Proposition 2.31. *If A is a circulant $n \times n$ matrix with first row $a_0, a_1, \ldots, a_{n-1}$, say, then the eigenvalues λ_j are given by*

$$\lambda_j := \sum_{k=0}^{n-1} a_k \omega_j^k \qquad (j = 0, 1, \ldots, n-1),$$

with corresponding eigenvectors $v_j := (1, \omega_j, \omega_j^2, \ldots, \omega_j^{n-1})^t$, where $\omega_j :=$ $\exp(\frac{2\pi i j}{n})$.

A class of graphs which are interesting with respect to arithmetic consists of the so-called *integral* graphs.

Definition 2.32. A graph $\mathcal{G}$ is called *integral* if $\mathrm{Spec}(\mathcal{G}) \subset \mathbb{Z}$.

In 1974 Frank Harary and Allen Schwenk introduced this concept and published a volume of Springer Lecture Notes entitled *"Which graphs have integral spectra?"* [16]. A graph theoretical classification of all integral graphs is still unknown. An important subclass of the integral graphs is considered in the following theorem and is thus much better understood than integral graphs in general.

Theorem 2.33 (Wasin So [46], 2005). *Every integral circulant graph $\mathcal{G}$ is uniquely characterized by a so-called gcd graph $\gcd(n,\mathcal{D}) = \mathrm{ICG}(n,\mathcal{D})$ defined as follows: n is the order of $\mathcal{G}$ and $\gcd(n,\mathcal{D}) = (\mathbb{Z}_n, E)$ is a Cayley graph on $\mathbb{Z}_n$, where $\mathcal{D}$ is a set of positive divisors of n and $E := \{(a,b) : a,b \in \mathbb{Z}_n,\ (a-b,n) \in \mathcal{D}\}$.*

Sketch of proof. Let $\mathcal{G}$ be any circulant graph. Circulant graphs are obviously regular, more precisely they are Cayley graphs on cyclic groups, i.e., w.l.o.g. $\mathcal{G} \simeq \mathrm{Cay}(\mathbb{Z}_n, S)$ for the order n of $\mathcal{G}$ and a suitable set S. By Proposition 2.31 we know that each eigenvalue $\lambda \in \mathrm{Spec}(\mathcal{G})$ is an integral linear combination of n-th roots of unity. By a famous result of Lam and Leung [26] published in 2000, one has a clear understanding of when sums of roots of unity vanish, and one could apply this result here. However, it luckily turns out in our situation that if S is a union of sets $S_d := \{d, 2d, \ldots, (\frac{n}{d}-1)d\}$ for some divisors $d \mid n$ satisfying $(d, \frac{n}{d}) = 1$, then each eigenvalue λ can be expressed as a sum of Ramanujan sums, a special sum well known in analytic number theory and, in particular, always with **integral values**. Consequently, we have $\mathrm{Spec}(\mathcal{G}) \subset \mathbb{Z}$. Of course, it remains to check that $\mathrm{Spec}(\mathcal{G}) \subset \mathbb{Z}$ imposes the above mentioned structure on S. It is then not difficult to check that $S = \bigcup S_d$ yields the asserted edge set E. □

In Section 4 we shall study arithmetical properties of integral circulant graphs.

Over the last decades the theory of *expander graphs* has attracted quite a lot of interest. These graphs with strong connectivity properties have a lot of applicatory consequences, e.g., the resolution of an extremal problem in communication network theory (cf. [6]), and they are also of importance in theoretical computer science (cf.[19]). A special class of expanders are *Ramanujan graphs*, which were introduced by Lubotzky, Phillips, and Sarnak [31] in 1988.

Definition 2.34. A k-regular graph $\mathcal{G} = (V,E)$ is called a *Ramanujan graph* or simply *Ramanujan*, if $\Lambda(\mathcal{G}) := \max\{|\lambda| \colon\ \lambda \in \mathrm{Spec}(\mathcal{G}),\ |\lambda| < k\}$ satisfies $\Lambda(\mathcal{G}) \leq 2\sqrt{k-1}$.

As we shall see in Theorem 3.20 below, these graphs are intimately linked with the theory of primes on graphs (see [35]). In Section 4 we shall examine which integral circulant graphs are Ramanujan graphs.

Exercise 2.35. Which of the cycle graphs $\mathcal{C}_n$ and the complete graphs $\mathcal{K}_n$ are Ramanujan?

3 The prime number theorem for graphs

We denote by $\mathbb{P}$ the set of prime numbers and, as usual, by

$$\pi(x) := \sum_{p \leq x,\ p \in \mathbb{P}} 1$$

the function counting the prime numbers up to x in the set of positive integers. There is indeed a concept of primality in graphs, and this will be defined for arbitrary finite, undirected and connected graphs $\mathcal{G}$. We introduce a corresponding prime counting function $\pi_{\mathcal{G}}(x)$ and relate it to the *Ihara zeta function*. This reveals analogies to other zeta functions, in particular the classical Riemann zeta function with its application to $\pi(x)$ (cf. [4], [23] or [37]). We shall derive determinant formulae and functional equations and consider a Riemann hypothesis, which is related to the *Ramanujan* property of a graph (cf. Def. 2.34). Finally, the Ihara zeta function is used to prove a graph-theoretic prime number theorem. Most of the material covered in this section is taken from Audrey Terras' book "Zeta Functions of Graphs – A Stroll through the Garden" [48].

3.1 Primes in friendly graphs

Assumption. The graphs $\mathcal{G} = (V, E)$ we consider always are **finite**, **undirected**, and **connected**.

Unless otherwise stated, we also assume that $\mathcal{G}$ has **no leaves** (= vertices of degree 1) and is **not a cycle graph** with or without leaves (i.e., $\mathcal{G} \setminus \{\text{leaves}\} \neq \mathcal{C}_n$ for all n, in other words $\mathcal{G}$ is not *unicyclic*).

A graph is called *friendly* if it satisfies the above assumption.

Exercise 3.1. Show that every friendly graph with at least one edge contains a cycle and, therefore, is *multicyclic*.

Let $\mathcal{G} = (V, E)$ be a friendly graph with $|E| = m$. We orient the m edges of $\mathcal{G}$ arbitrarily to obtain directed edges $e_1, e_2, \ldots, e_m$ (with arbitrary labelling). Then we label the *inverse* edges by setting $e_{m+j} := e_j^{-1}$, where $e_j^{-1} = (v, u)$ for $e_j = (u, v)$. The graph $\vec{\mathcal{G}} = (V, \vec{E})$ with $\vec{E} := E \cup E^{-1}$ is called an *orientation* of $\mathcal{G}$.

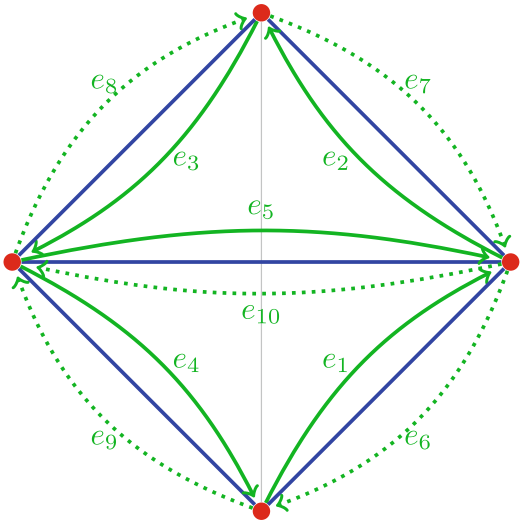


Figure 4.8. An orientation of $\mathcal{K}_4 - e$

Example 3.2 (An orientation of $\mathcal{K}_4 - e$).

Definition 3.3. Let $\mathcal{G} = (V, E)$ be a graph, and let $\vec{\mathcal{G}} = (V, \vec{E})$ be an orientation of $\mathcal{G}$. For $m := |E| = \frac{1}{2}|\vec{E}|$, the $2m \times 2m$ matrix $W_{\vec{\mathcal{G}}} = (w_{ij})$ is defined as follows: Set $w_{ij} = 1$ if $e_i e_j$ is a directed path (of length 2) in $\vec{\mathcal{G}}$, but not $e_i^{-1} = e_j$, and otherwise $w_{ij} := 0$. Then $W_{\vec{\mathcal{G}}}$ is called the *edge adjacency matrix* of $\vec{\mathcal{G}}$.

Definition 3.4. Let $C = e_1 e_2 \ldots e_s$ be a path of edges in an orientation $\vec{\mathcal{G}}$ of a graph $\mathcal{G}$.

- C is said to have a *backtrack* if $e_{j+1} = e_j^{-1}$ for some $j \in \{1, 2, \ldots, s-1\}$ or $e_s = e_1^{-1}$.
- If C is closed but not the power of a shorter cycle, then we call it a *primitive* or *prime path* if it has no backtracks.
- Two cycles are said to be *equivalent* if we get one from the other just by changing the starting vertex. For C closed,

$$[C] := \{C,\ e_2 e_3 \ldots e_s e_1,\ e_3 \ldots e_s e_1 e_2,\ \ldots\ ,\ e_s e_1 e_2 \ldots e_{s-1}\}$$

  is called the *equivalence class* of C.
- A *prime* in $\mathcal{G}$ is an equivalence class $[P]$ of prime paths, i.e., P is a primitive path.
- $\nu[P] := \nu(P)$ is called the *length of the prime* $[P]$.

Example 3.5.

(i) Primes in $\mathcal{K}_4 - e$ are (Ex. 3.2 cont'd) :

- $[C_1]$ for $C_1 := e_2e_3e_5$ and $[e_{10}e_8e_7] = [C_1^{-1}]$ with lengths $\nu[C_1] = \nu[C_1^{-1}] = 3$;
- $[C_2]$ for $C_2 := e_1e_2e_3e_4$ with length $\nu[C_2] = 4$;
- $[C_2^nC_3]$ for $C_3 := e_1e_{10}e_4$ and each $n \geq 0$, hence there are infinitely many primes in $\mathcal{K}_4 - e$.

(ii) In naughty graphs one could also look for primes, but the example at the beginning of this section shows that there are only two primes, namely the two orientations of the unique cycle.

Exercise 3.6.

(i) Check that equivalence between cycles is in fact an equivalence relation.

(ii) Let C be a prime path. Show that each closed path equivalent with C is also a prime path.

(iii) Verify that powers of primes are the only non-primes among cycles.

(iv) If C is a cycle, we can *factorize it into subcycles* C_1 and C_2 if C_1 and C_2 have a common vertex v and, starting from v, we obtain C by concatenation of C_1 and C_2. Give examples of the fact that we do not have unique factorization of cycles into primes, e.g., in Example 3.5.

Definition 3.7. Let $\mathcal{G}$ be a friendly graph, and let $\mathbb{P}_\mathcal{G}$ be the set of primes in $\mathcal{G}$. Then

$$\pi_\mathcal{G}(k) := \#\{[P] \in \mathbb{P}_\mathcal{G} : \ \nu[P] = k\}$$

is called the *prime counting function*. (Observe the $=$ sign as opposed to $\leq$ in the prime counting function for integers).

An important parameter will be the **greatest common divisor of the prime path lengths**

$$\Delta_\mathcal{G} := \gcd\{\nu[P] : \ [P] \in \mathbb{P}_\mathcal{G}\} \tag{4.2}$$

in a friendly graph $\mathcal{G}$.

Proposition 3.8. *Let $\mathcal{G}$ be a friendly graph. Then $\pi_\mathcal{G}(k) = 0$ if $\Delta_\mathcal{G} \nmid k$.*

Proof. Since $\Delta_\mathcal{G} \mid \nu[P]$ for each prime $[P]$, there is no prime $[P]$ with $\nu[P] = k$ for $\Delta_\mathcal{G} \nmid k$. □

3.2 Ihara's zeta function The **Riemann zeta function** encodes information about the **distribution of primes** in the set of positive integers and can be used to prove the **prime number theorem** (cf. [4], [23] or [37]). Other zeta functions have been defined and studied in order to obtain information about mathematical objects like prime numbers in arithmetic progressions (**Dirichlet L-functions**), prime ideals in rings of algebraic integers (**Dedekind zeta functions**), primitive closed geodesics in manifolds (**Selberg zeta function**), and many more (see [24], [27], [33], [38]). We shall look at a zeta function related to primes in graphs and will uncover analogies with the other zeta functions.

The Ihara zeta function was first defined by Yasutaka Ihara [21] in the 1960s in the context of discrete subgroups of the two-by-two p-adic special linear group. Jean-Pierre Serre suggested 1977 in his book *"Arbres, Amalgames, SL_2"* (=*"Trees"* [43]) that Ihara's original definition could be reinterpreted graph-theoretically, namely as a zeta function related to closed geodesics in graphs (compare with the Selberg zeta function). In 1985 Toshikazu Sunada [47] put this suggestion into practice.

Definition 3.9. Let $\mathcal{G}$ be a friendly graph. The *Ihara zeta function* is the complex function

$$\zeta_{\mathcal{G}}(u) = \zeta(u, \mathcal{G}) := \prod_{[P]\in\mathbb{P}_{\mathcal{G}}} \left(1 - u^{\nu[P]}\right)^{-1}, \tag{4.3}$$

where $u \in \mathbb{C}$ is supposed to have sufficiently small absolute value. Recall that we distinguish between the primes $[P]$ and $[P^{-1}]$. The question of convergence will be postponed until after Proposition 3.11.

Example 3.10. Although the cycle graph $\mathcal{C}_n$ is not friendly (with only two primes = cycle in two orientations), we still can evaluate its Ihara zeta function:

$$\zeta_{\mathcal{C}_n}(u) = (1 - u^n)^{-2} = \left(\sum_{k=0}^{\infty} u^{kn}\right)^2.$$

Let us explain at the outset why the Ihara zeta function of a friendly graph $\mathcal{G}$ is useful in dealing with primes in $\mathcal{G}$. To this end, let $N_m(\mathcal{G})$ be the number of cycles of length m without backtracks in $\mathcal{G}$.

Proposition 3.11. *For all $u \in \mathbb{C}$ with sufficiently small absolute value $|u|$ it holds that*

$$\log \zeta_{\mathcal{G}}(u) = \sum_{m=1}^{\infty} \frac{N_m(\mathcal{G})}{m} u^m. \tag{4.4}$$

Proof. Taking logarithms in (4.3) we obtain

$$\begin{aligned}\log \zeta_{\mathcal{G}}(u) &= \log \prod_{[P]\in\mathbb{P}_{\mathcal{G}}} \left(1 - u^{\nu[P]}\right)^{-1} = -\sum_{[P]\in\mathbb{P}_{\mathcal{G}}} \log\left(1 - u^{\nu[P]}\right) \\ &= \sum_{[P]\in\mathbb{P}_{\mathcal{G}}} \sum_{j=1}^{\infty} \frac{1}{j} u^{j\nu[P]} = \sum_{P \text{ primitive}} \frac{1}{\nu(P)} \sum_{j=1}^{\infty} \frac{1}{j} u^{j\nu(P)} \\ &= \sum_{P \text{ primitive}} \sum_{j=1}^{\infty} \frac{1}{\nu(P^j)} u^{\nu(P^j)},\end{aligned}$$

since there are exactly $\nu(P)$ elements in a prime $[P]$ and trivially $\nu(P^j) = j\nu(P)$. As seen in Exercise 3.6 (iii), any cycle without backtrack in $\mathcal{G}$ is a power of some primitive cycle. Hence

$$\begin{aligned}\log \zeta_{\mathcal{G}}(u) &= \sum_{C \text{ cycle w/o backtr.}} \frac{1}{\nu(C)} u^{\nu(C)} \\ &= \sum_{m=1}^{\infty} \frac{N_m(\mathcal{G})}{m} u^m. \end{aligned}\qquad \square$$

Now let us verify that $\zeta_{\mathcal{G}}(u)$ does indeed converge inside some circle in the complex plane with positive radius centered at 0. By Proposition 2.24, $(A_{\mathcal{G}}^m)_{jj}$ equals the number of directed paths of length m from $v_j \in V$ to v_j, which in turn is the number of directed cycles of length m containing v_j. Hence $(A_{\mathcal{G}}^m)_{jj}$ is greater than or equal to the number of cycles of length m containing v_j. Therefore, setting $n := |V|$ for the number of vertices in $\mathcal{G}$, we have

$$\sum_{j=1}^{n} (A_{\mathcal{G}}^m)_{jj} \geq N_m(\mathcal{G}).$$

Since $A_{\mathcal{G}}$ has only non-negative entries 0 and 1, it is easy to see that

$$(A_{\mathcal{G}}^m)_{ij} \leq (\mathbb{1}^m)_{ij},$$

where $\mathbb{1}$ denotes the $n \times n$-matrix with all entries 1. Obviously, $\mathbb{1}^m = n^{m-1}\mathbb{1}$, hence

$$N_m(\mathcal{G}) \leq \sum_{j=}^{n} n^{m-1} = n^m.$$

For $|u| < \frac{1}{n}$ the sum

$$\sum_{m=1}^{\infty} \frac{N_m(\mathcal{G})}{m} u^m$$

converges absolutely. By Proposition 3.11, this implies that $\log \zeta_{\mathcal{G}}(u)$ is well defined for $|u| < \frac{1}{n}$, and thus the same is true for $\zeta_{\mathcal{G}}(u)$ itself. □

We introduce the notion of the *radius of convergence* $R_{\mathcal{G}}$ such that $\zeta_{\mathcal{G}}(u)$ converges inside $|u| < R_{\mathcal{G}}$ and has a singularity on the border $|u| = R_{\mathcal{G}}$, called *circle of convergence*.

One might like to compare formula (4.4) with related identities for the classical **Riemann zeta function** and the **von Mangoldt function** or other characteristic functions for primes in $\mathbb{Z}$ (cf. [4] or [23]).

We are lucky to have explicit formulae which allow us to compute $\zeta_{\mathcal{G}}(u)$. One of them is called the *two-term determinant formula* (Theorem 3.31) and relates $\zeta_{\mathcal{G}}(u)$ to the determinant of the edge adjacency matrix $W_{\vec{\mathcal{G}}}$ (cf. Definition 3.3). This formula can be used to obtain the so-called *Ihara three-term determinant formula* (Theorem 3.15).

In order to formulate the latter result, we should consider the *fundamental group* of a graph and its rank. The fundamental group, in general associated to any given pointed topological space, provides a way to determine when two paths, starting and ending at a fixed base point, hence called loops, can be continuously deformed into each other. Such loops are considered equivalent. Two loops can be combined by just wandering around the first one and afterwards around the second one. The elements of the fundamental group are the equivalence classes of loops and the operation is the combination of these.

One of the simplest examples is the fundamental group of the circle, which is obviously isomorphic to the group $(\mathbb{Z}, +)$. This corresponds to the fundamental group of any of the cycle graphs $\mathcal{C}_n$ (with or without leaves). In general, it turns out that the fundamental group $\Gamma(\mathcal{G}, v_1)$ of a graph $\mathcal{G} = (V, E)$ relative to the base vertex $v_1 \in V$ is a *free group* on r generators, where r is the *rank* of the group. $\Gamma(\mathcal{G}, v_1)$ and, in particular, its rank r can easily be determined by the following

Algorithm 3.12. Let $\mathcal{G} = (V, E)$ be a graph (possibly directed with multi-edges and loops).

(1) Choose any orientation of $\mathcal{G}$ to obtain $\vec{\mathcal{G}}$.

(2) Choose any spanning tree $\mathcal{T} = (V, E')$ of $\mathcal{G}$ (which exists by Proposition 2.9).

(3) For each edge $v_i v_j \in E \setminus E'$, construct a path from v_1 to v_i and from v_j back to v_1 (all these paths are unique by Exercise 2.8) to obtain a cycle starting from v_1 via v_i and v_j back to v_1.

Proposition 3.13. *Let $\mathcal{G} = (V, E)$ be a graph with $v_1 \in V$.*

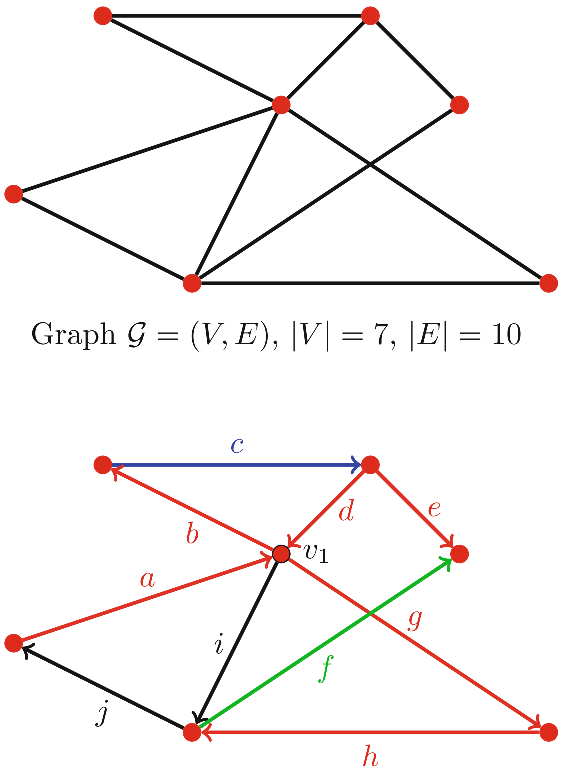
(i) *The equivalence classes of the cycles obtained in (3) of Algorithm 3.12 form a free basis of $\Gamma(\mathcal{G}, v_1)$.*

(ii) *The rank r of $\Gamma(\mathcal{G}, v_1)$ satisfies $r = |E| - |V| + 1$.*

Proof. The fundamental group $\Gamma(\mathcal{G}, v_1)$ consists of sequences of cycles (more precisely, equivalence classes of cycles) starting from v_1. The spanning tree constructed in step (2) of Algorithm 3.12 does not have any cycles. However, each edge in $E \setminus E'$ completes a unique cycle, and different edges in $E \setminus E'$ yield inequivalent cycles. This proves (i).

By (i) we know that $r = |E| - |E'|$, where E' is the edge set of a spanning tree of $\mathcal{G}$. By Exercise 2.8, a tree on $|V|$ vertices has $|V| - 1$ edges. This proves (ii). □

Example 3.14. We determine a free basis and the rank of the fundamental group of the graph $\mathcal{G}$ below:



Graph $\mathcal{G} = (V, E)$, $|V| = 7$, $|E| = 10$

Spanning tree + orientation of $\mathcal{G}$

c : bcd

f : $ghfe^{-1}d$

i : $ih^{-1}g^{-1}$

j : $ghja$

Fundamental group:
$\Gamma(\mathcal{G}, v_1)$ generated by $[bcd]$, $[ghfe^{-1}d]$, $[ih^{-1}g^{-1}]$ and $[ghja]$

Rank r of $\Gamma(\mathcal{G}, v_1)$:
$r = 4 = 10 - 7 + 1 = |E| - |V| + 1$

Figure 4.9. How to determine the fundamental group of a graph

(1) We first choose a vertex v_1, any orientation of $\mathcal{G}$, and any spanning tree $\mathcal{T}$ of $\mathcal{G}$.

(2) For each directed edge e in $\mathcal{G}$ not belonging to $\mathcal{T}$, we construct the unique directed cycle in $\mathcal{G}$ starting from v_1 through e.

(3) These cycles are a free basis of $\Gamma(\mathcal{G}, v_1)$ and their number is the rank of the fundamental group.

Now we are able to formulate the three-term determinant formula, first proved by Y. Ihara [21] in 1966 and later generalized by K. Hashimoto [18] in 1989 and H. Bass [5] in 1992.

Theorem 3.15 (Ihara's three-term determinant formula). *Let $\mathcal{G} = (V, E)$ be a friendly graph with $V = \{v_1, \ldots, v_n\}$. Let $Q_{\mathcal{G}} = (q_{ij})$ be the $n \times n$ diagonal matrix with diagonal entries $q_{jj} = d(v_j) - 1$, where $d(v_j)$ is the degree of v_j. Then*

$$\zeta_{\mathcal{G}}(u)^{-1} = (1 - u^2)^{r-1} \det(I - A_{\mathcal{G}} u + Q_{\mathcal{G}} u^2),$$

where I is the unit matrix and $r = |E| - |V| + 1$ is the rank of the fundamental group of $\mathcal{G}$.

Before proving Theorem 3.15 in Section 3.3, let us deduce a variety of consequences of the three-term determinant formula.

Example 3.16. (cf. Fig. 4.10 and 4.11) Consider the complete graph $\mathcal{K}_4$ with $r = 6 - 4 + 1 = 3$,

$$A_{\mathcal{K}_4} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

and $Q_{\mathcal{K}_4} = 2I$, because $\mathcal{K}_4$ is 3-regular. The three-term determinant formula yields

$$\zeta_{\mathcal{K}_4}(u)^{-1} = (1 - u^2)^2 \det \begin{pmatrix} 2u^2 + 1 & -u & -u & -u \\ -u & 2u^2 + 1 & -u & -u \\ -u & -u & 2u^2 + 1 & -u \\ -u & -u & -u & 2u^2 + 1 \end{pmatrix}$$

$$= (1 - u^2)^2 (1 - u)(1 - 2u)(1 + u + 2u^2)^3.$$

Hence the five poles of $\zeta_{\mathcal{K}_4}(u)$ with different multiplicities are located as follows: three real poles at ±1 and $\frac{1}{2}$, and two complex poles at $\frac{1}{4}(-1 \pm i\sqrt{7})$ (with absolute value $\frac{\sqrt{2}}{2}$). Since the pole closest to the origin lies at $\frac{1}{2}$, we have $R_{\mathcal{K}_4} = \frac{1}{2}$ as radius of convergence.

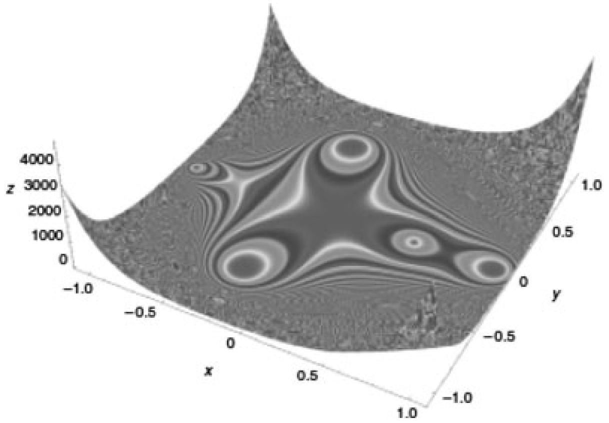


Figure 4.10. Ihara's zeta function for $\mathcal{K}_4$ in the u-variable: $z = |\zeta_{\mathcal{K}_4}(x+iy)|^{-1}$. This figure is taken from [48] with permission of the author.

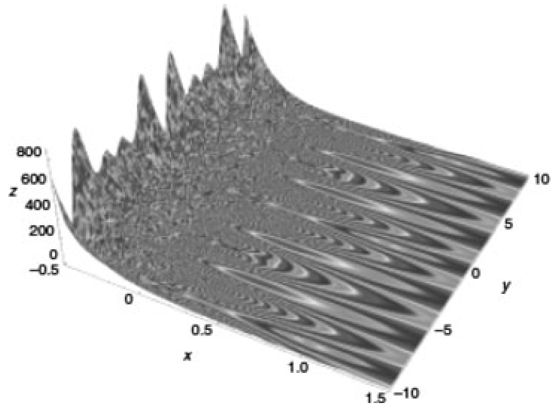


Figure 4.11. Ihara's zeta function for $\mathcal{K}_4$ in the s-variable: $z = |\zeta_{\mathcal{K}_4}(2^{-(x+iy)})|^{-1}$. This figure is taken from [48] with permission of the author.

Exercise 3.17. Show that the irregular graph $\mathcal{K}_4 - e$ (cf. Example 3.2) satisfies

$$\zeta_{\mathcal{K}_4 - e}(u)^{-1} = (1-u^2)(1-u)(1+u^2)(1+u+2u^2)(1-u^2-2u^3)$$

with nine roots: ± 1, $\pm i$, $\frac{1}{4}(-1 \pm i\sqrt{7})$ and three cubic roots u_1, u_2, u_3, say, satisfying $|s_1| \approx 0.657$ and $|s_2| = |s_3| \approx 0.872$. Hence the radius of convergence is $R_{\mathcal{K}_4 - e} \approx 0.657$.

We shall now specialize on regular graphs. For this graph class a lot of the concepts we are interested in are partly easier to understand, but most results can be generalized to irregular graphs.

Reconsider the regular graph $\mathcal{K}_4$ (cf. Example 3.16) with

$$\zeta_{\mathcal{K}_4}(u)^{-1} = (1-u^2)^2(1-u)(1-2u)(1+u+2u^2)^3.$$

We take a look at $z = |\zeta_{\mathcal{K}_4}(x+iy)|^{-1}$ (Fig. 4.10), depicting the five zeros (not counting multiplicities).

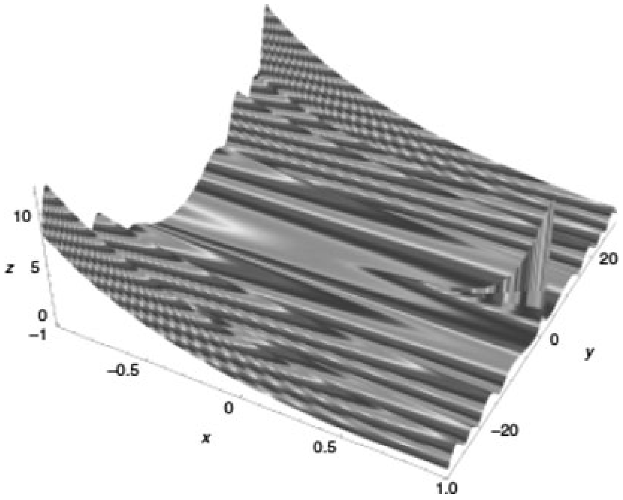


Figure 4.12. Riemann's zeta function $z = |\zeta(x + iy)|$. This figure is taken from [48] with permission of the author.

To catch another (logarithmic) sight of $z = |\zeta_{\mathcal{K}_4}(u)|^{-1}$, we make the change of complex variable $u \mapsto \frac{1}{2^u}$ and display the landscape of $z = |\zeta_{\mathcal{K}_4}(2^{-(x+iy)})|^{-1}$ (Fig. 4.11).

This bears some resemblance to the landscape $z = |\zeta(x + iy)|$ of Riemann's zeta function $\zeta(s)$ (Fig. 4.12).

In fact, the resemblance between the zeta functions of Riemann and Ihara is striking. The classical **Riemann hypothesis** (see e.g. [23] or [37]) conjectures the location of the non-trivial zeros of the Riemann zeta function.

Definition 3.18. Let $\mathcal{G}$ be a friendly $(q+1)$-regular graph and consider the Ihara zeta function $\zeta_{\mathcal{G}}(q^{-s})$ as a complex function in $s \in \mathbb{C}$. We say that $\zeta_{\mathcal{G}}(q^{-s})$ satisfies the *Riemann hypothesis* if $\zeta_{\mathcal{G}}(q^{-s})^{-1}$ can vanish in the critical strip $0 < \operatorname{Re} s < 1$ only for $\operatorname{Re} s = \frac{1}{2}$, i.e., for $|u| = |q^{-s}| = \frac{1}{\sqrt{q}}$.

Exercise 3.19. Check that $\zeta_{\mathcal{K}_4}(2^{-s})$ satisfies the Riemann hypothesis (cf. Example 3.16).

Theorem 3.20. *Let $\mathcal{G}$ be a friendly $(q+1)$-regular graph. Then $\zeta_{\mathcal{G}}(q^{-s})$ satisfies the Riemann hypothesis if and only if $\mathcal{G}$ is a Ramanujan graph (cf. Definition 2.34).*

Proof. By the determinant formula in Theorem 3.15, we have

$$\zeta_{\mathcal{G}}(q^{-s})^{-1} = (1 - q^{-2s})^{r-1} \det(I - A_{\mathcal{G}} q^{-s} + Q_{\mathcal{G}} q^{-2s}). \qquad (4.5)$$

Since $A_{\mathcal{G}}$ is a real symmetric matrix, it is diagonalizable (even by means of orthonormal matrices). Hence there is an invertible matrix T such that $T^{-1} A_{\mathcal{G}} T = L_{\mathcal{G}}$, where

$L_{\mathcal{G}}$ is a diagonal matrix with the eigenvalues of $\mathcal{G}$ on the diagonal. Since $\mathcal{G}$ is assumed to be $(q+1)$-regular, we have $Q_{\mathcal{G}} = qI$. Altogether

$$\begin{aligned} \det(I - A_{\mathcal{G}} q^{-s} + Q_{\mathcal{G}} q^{-2s}) &= \det\Big(TIT^{-1} - T(q^{-s}L_{\mathcal{G}})T^{-1} + T(q^{1-2s}I)T^{-1}\Big) \\ &= \det T \cdot \det\Big(I - q^{-s}L_{\mathcal{G}} + q^{1-2s}I\Big) \cdot \det T^{-1} \\ &= \det\Big(I - q^{-s}L_{\mathcal{G}} + q^{1-2s}I\Big) \\ &= \prod_{\lambda \in \mathrm{Spec}(\mathcal{G})} \big(1 - \lambda q^{-s} + q^{1-2s}\big), \end{aligned}$$

hence

$$\zeta_{\mathcal{G}}(q^{-s})^{-1} = (1 - q^{-2s})^{r-1} \prod_{\lambda \in \mathrm{Spec}(\mathcal{G})} \big(1 - \lambda q^{-s} + q^{1-2s}\big).$$

Fix some $\lambda \in \mathrm{Spec}(\mathcal{G})$ and write

$$1 - \lambda q^{-s} + q^{1-2s} = (1 - \alpha_1 q^{-s})(1 - \alpha_2 q^{-s}),$$

i.e., α_1 and α_2 are reciprocals of poles of $\zeta_{\mathcal{G}}(q^{-s})$ satisfying $\alpha_1 + \alpha_2 = \lambda$ and $\alpha_1\alpha_2 = q$. This implies the quadratic equation $\alpha_1^2 - \lambda\alpha_1 = -q$, leading to

$$\alpha_1 = \frac{1}{2}\big(\lambda + \sqrt{\lambda^2 - 4q}\,\big) \quad \text{and} \quad \alpha_2 = \frac{1}{2}\big(\lambda - \sqrt{\lambda^2 - 4q}\,\big), \tag{4.6}$$

or vice versa.

By Proposition 2.26 we know that $q + 1 \in \mathrm{Spec}(\mathcal{G})$ and $|\lambda| \leq q + 1$ for all $\lambda \in \mathrm{Spec}(\mathcal{G})$. We distiguish three cases, namely $\lambda = \pm(q+1)$, $|\lambda| \leq 2\sqrt{q}$, and $2\sqrt{q} < |\lambda| < q + 1$.

Case 1: $\lambda = \pm(q+1)$.

By (4.6), we immediately obtain $\alpha_1 = \pm q$ and $\alpha_2 = \pm 1$, or vice versa.

Case 2: $|\lambda| \leq 2\sqrt{q}$.

It follows that $\lambda^2 - 4q \leq 0$, thus $\alpha_1, \alpha_2 \in \mathbb{C}$ and $\sqrt{\lambda^2 - 4q} = \pm i \cdot \sqrt{4q - \lambda^2}$. This implies that

$$|\alpha_1| = \frac{1}{2}\big|\lambda \pm i \cdot \sqrt{4q - \lambda^2}\big| = \frac{1}{2}\sqrt{\lambda^2 + (4q - \lambda^2)} = \sqrt{q},$$

and similarly $|\alpha_2| = \sqrt{q}$.

Case 3: $2\sqrt{q} < |\lambda| < q + 1$.

Now $\alpha_1, \alpha_2 \in \mathbb{R}$. Observing that $\lambda + \sqrt{\lambda^2 - 4q}$ is increasing in λ, we obtain by (4.6) for positive λ, i.e., for $2\sqrt{q} < \lambda < q + 1$, that $\sqrt{q} < \alpha_1 < q$. Since $\alpha_1\alpha_2 = q$, this implies $1 < \alpha_2 < \sqrt{q}$. Checking the vice versas and $\lambda < 0$, we conclude that $1 < |\alpha_i| < q$, but $|\alpha_i| \neq \sqrt{q}$ for $i = 1, 2$.

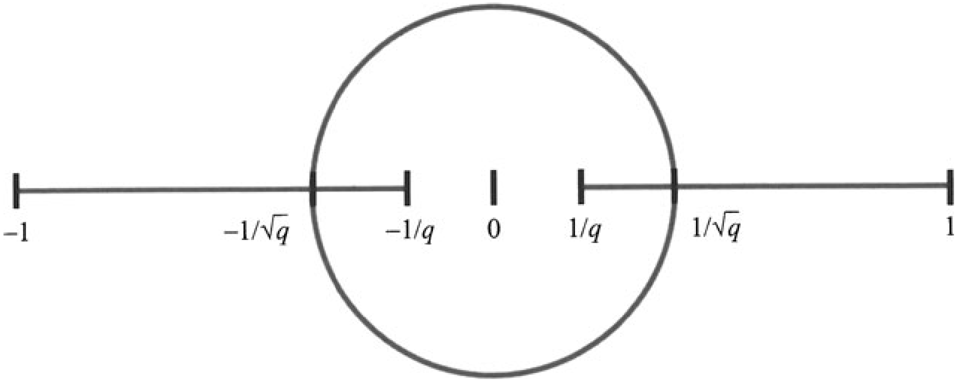


Figure 4.13. Possible locations for poles of $\zeta_{\mathcal{G}}(u)$ for regular $\mathcal{G}$. This figure is taken from [48] with permission of the author.

To complete the proof we just have to remember that the poles of $\zeta_{\mathcal{G}}(q^{-s})$, other than those coming from the first factor in (4.5) and having $\Re s = 0$, are the reciprocals of some α_i, i.e., $\alpha_i = q^s$ and $|\alpha_i| = |q^{\Re s + i\Im s}| = q^{\Re s}$. The eigenvalues $\pm(q+1) \in \operatorname{Spec}(\mathcal{G})$ are maximal in absolute value and therefore irrelevant for the Ramanujan property. At the same time, they do not affect the Riemann hypothesis, because they correspond to poles having $\Re s = 1$ or $\Re s = 0$ (see Case 1). The eigenvalues considered in Case 2 satisfy the Ramanujan condition and the corresponding α_i all have $\Re s = \frac{1}{2}$, satisfying the Riemann hypothesis, as shown in Case 2. If there existed an eigenvalue $\lambda \in \operatorname{Spec}(\mathcal{G})$ belonging to Case 3, it would violate the Ramanujan condition. At the same time, Case 3 has revealed the existence of a pole with $0 < \Re s < 1$, but $\Re s \neq \frac{1}{2}$, thus contradicting the Riemann hypothesis. □

The preceding proof immediately implies

Corollary 3.21. *Let $\mathcal{G}$ be a friendly $(q+1)$-regular graph.*

(i) *The figure above shows the possible locations for the poles of $\zeta_{\mathcal{G}}(u)$. Poles satisfying the Riemann hypothesis are those on the circle.*

(ii) *The radius of convergence satisfies $R_{\mathcal{G}} = \frac{1}{q}$.*

Setting again $u = q^{-s}$, there are several functional equations for the Ihara zeta function on a regular graph, relating the value at s to that at $1-s$ (corresponding to $u \leftrightarrow \frac{1}{qu}$), just as in the case of the Riemann zeta function.

Theorem 3.22 (Functional equations for Ihara zeta functions). *Let $\mathcal{G}$ be friendly $(q+1)$-regular graph of order n and rank r of its fundamental group. Then, among other similar identities, we have*

(i) $\xi^{(1)}_{\mathcal{G}}(u) := (1-u^2)^{r-1+\frac{n}{2}}(1-q^2u^2)^{\frac{n}{2}}\zeta_{\mathcal{G}}(u) = (-1)^n\xi^{(1)}_{\mathcal{G}}(\frac{1}{qu})$;

(ii) $\xi^{(2)}_{\mathcal{G}}(u) := (1+u)^{r-1}(1-u)^{r-1+n}(1-qu)^n\zeta_{\mathcal{G}}(u) = \xi^{(2)}_{\mathcal{G}}(\frac{1}{qu})$;

(iii) $\xi^{(3)}_{\mathcal{G}}(u) := (1-u^2)^{r-1}(1+qu)^n\zeta_{\mathcal{G}}(u) = \xi^{(3)}_{\mathcal{G}}(\frac{1}{qu})$.

Proof of (i): The determinant formula in Theorem 3.15 and the transformation $u \mapsto \frac{1}{qu}$ imply

$$\begin{aligned}\xi_{\mathcal{G}}^{(1)}(u) &= (1-u^2)^{\frac{n}{2}}(1-q^2u^2)^{\frac{n}{2}} \det(I - A_{\mathcal{G}}u + qu^2 I)^{-1} \\ &= \left(\frac{q^2}{q^2u^2} - 1\right)^{\frac{n}{2}} \left(\frac{1}{q^2u^2} - 1\right)^{\frac{n}{2}} \det\left(I - \frac{1}{qu}A_{\mathcal{G}} + \frac{q}{q^2u^2} I\right)^{-1} \\ &= (-1)^n \xi_{\mathcal{G}}^{(1)}\left(\frac{1}{qu}\right).\end{aligned}$$ □

Exercise 3.23.

(i) Check (ii) and (iii) in Theorem 3.22.

(ii) Show that $\zeta_{\mathcal{G}}$ has a pole at $\frac{1}{qu}$ if it has one at u. Verify that $\frac{1}{qu}$ is the complex conjugate of u if u lies on the circle of radius $\frac{1}{\sqrt{q}}$, and that $\frac{1}{qu}$ lies in the interval $(\frac{1}{q}, \frac{1}{\sqrt{q}})$ for $u \in (\frac{1}{\sqrt{q}}, 1)$.

Most of the results in this section can be generalized to irregular graphs. All we shall present here is a theorem showing the possible location of the poles of $\zeta_{\mathcal{G}}$ for irregular $\mathcal{G}$.

Theorem 3.24 (Motoko Kotani and Toshikazu Sunada [25], 2000). *Let $\mathcal{G} = (V, E)$ be a friendly graph with $p + 1 := \min\{d(v) :\ v \in V\}$ and $q + 1 := \max\{d(v) : v \in V\}$, and let $R_{\mathcal{G}}$ be the radius of convergence of $\zeta_{\mathcal{G}}$.*

(i) *We have $\frac{1}{q} \le R_{\mathcal{G}} \le \frac{1}{p}$, and every pole u of $\zeta_{\mathcal{G}}(u)$ satisfies $R_{\mathcal{G}} \le |u| \le 1$.*

(ii) *Every non-real pole $u \in \mathbb{C}$ of $\zeta_{\mathcal{G}}(u)$ lies in the ring $\frac{1}{\sqrt{q}} \le |u| \le \frac{1}{\sqrt{p}}$.*

(iii) *The poles u of $\zeta_{\mathcal{G}}(u)$ lying on the circle $|u| = R_{\mathcal{G}}$ have the form*

$$u = R_{\mathcal{G}} e^{\frac{2\pi i k}{\Delta_{\mathcal{G}}}} \qquad (k = 1, 2, \ldots, \Delta_{\mathcal{G}}), \tag{4.7}$$

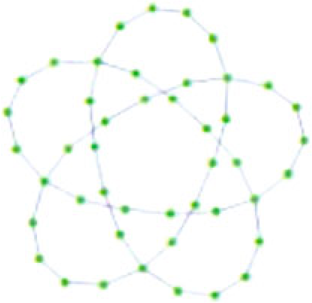
where $\Delta_{\mathcal{G}} = \gcd\{\nu[P] :\ [P] \in \mathbb{P}_{\mathcal{G}}\}$, as defined in (4.2).

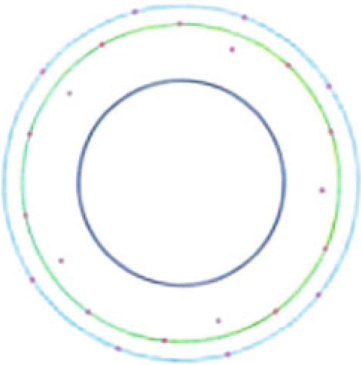
The most interesting part of the preceding theorem is the characterization in (iii). It is a rather straightforward consequence of an advanced result in linear algebra called the *Perron–Frobenius theorem* (cf. [20]), which essentially states:

> *Let M be an $n \times n$ matrix with non-negative real entries and* irreducible, *i.e. $(I + M)^{n-1}$ has only positive entries. Then the eigenvalues $\lambda_k \in$* Spec(M) *$(k = 1, \ldots, K)$, say, of maximal absolute value are given by* $\lambda_k = \max \operatorname{Spec}(M) \cdot e^{\frac{2\pi i k}{K}}$.

Applying this to the edge adjacency matrix $M := W_{\vec{\mathcal{G}}}$ proves (iii).

Example 3.25. The irregular graph $\mathcal{G}$ on the left below is obtained by adding four vertices to each edge of the complete graph $\mathcal{K}_5$. Clearly, $p + 1 = 2$ and $q + 1 = 4$.





On the right, all poles $u \neq \pm 1$ of $\zeta_{\mathcal{G}}(u)$ are depicted (observe 5-fold rotational symmetry). The circles centered at 0 have inverse radii $\frac{1}{\sqrt{q}} = \frac{\sqrt{3}}{3} \approx 0.577$, $\sqrt{R_{\mathcal{G}}} = \sqrt{\sqrt[5]{\frac{1}{3}}} \approx 0.896$, and $\frac{1}{\sqrt{p}} = 1$, coming from $\zeta_{\mathcal{G}}(u)^{-1} = \zeta_{\mathcal{K}_5}^{-1}(u^5) = (1 - u^{10})^5 (1 - 3u^5)(1 - u^5)(1 + u^5 + 3u^{10})$. (The figures are taken from [48] with permission of the author.)

Exercise 3.26. Prove that in our case $K = \Delta_{\mathcal{G}}$ (cf. [48], p. 95).

We shall see in Exercise 3.34 how to produce pictures like Figure 4.14.

3.3 The determinant formulae for Ihara's zeta function

A proof of Ihara's three-term determinant formula has still to be given. We shall proceed by first showing the two-term determinant formula (Theorem 3.31) mentioned above and derive Theorem 3.15 from it. First we have to introduce the exponential function and the logarithmic function for square matrices (as one usually does in the theory of Lie groups).

Definition 3.27. For an $n \times n$ matrix X over $\mathbb{R}$ or $\mathbb{C}$, the matrix

$$\exp X := \sum_{k=0}^{\infty} \frac{1}{k!} X^k$$

is called the *matrix exponential* of X.

Proposition 3.28. *Let X, Y be $n \times n$ matrices over $\mathbb{R}$ or $\mathbb{C}$.*

(i) *The series defining* $\exp X$ *converges absolutely.*

(ii) *For the zero matrix O we have* $\exp\ O = I$.

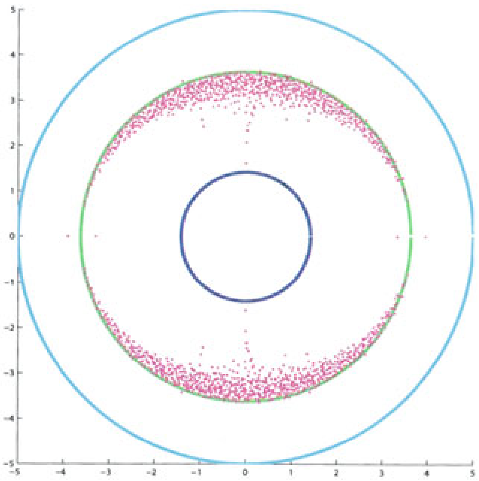


Figure 4.14. Locations of poles of $\zeta_{\mathcal{G}}(u)$ for an irregular random graph $\mathcal{G}$ with 800 vertices:
– violet circle of radius $\sqrt{p}$, blue circle of radius $\sqrt{q}$;
– green (Riemann hypothesis) circle of radius $\frac{1}{\sqrt{R_{\mathcal{G}}}}$.
RH says spectrum should lie inside green circle – almost true!
This figure is taken from [48] with permission of the author.

(iii) *If* $X = (x_{ij})$ *is diagonal, then* $\exp X = (\tilde{x}_{ij})$ *with* $\tilde{x}_{ii} = e^{x_{ii}}$ *and* $\tilde{x}_{ij} = 0$ *for* $i \neq j$.

(iv) *If* $XY = YX$, *then* $(\exp X)(\exp Y) = (\exp Y)(\exp X) = \exp(X + Y)$.

(v) *If* Y *is invertible, then* $\exp(Y^{-1}XY) = Y^{-1}(\exp X)Y$.

(vi) $\det(\exp X) = e^{\operatorname{tr} X}$.

Proof of (vi). For any X there is an invertible matrix Y such that YXY^{-1} is an upper triangular matrix T, say (see any book on linear algebra). Hence

$$\begin{aligned}\det(\exp X) &= \det(\exp(Y^{-1}TY)) \overset{(v)}{=} \det(Y^{-1} \cdot (\exp T) \cdot Y) \\ &= \det(Y^{-1}) \cdot \det(\exp T) \cdot \det Y = \det(\exp T).\end{aligned}$$

Since $T = (t_{ij})$ is upper triangular, so is $T^k = (t_{ij}^{(k)})$ for all k. In particular, $t_{ii}^{(k)} = t_{ii}^k$ for all i and k. This implies, by the definition of the matrix exponential, that $\exp T = (\tilde{t}_{ij})$ is upper diagonal with $\tilde{t}_{ii} = e^{t_{ii}}$ for all i. Hence

$$\det(\exp T) = e^{t_{11}} \cdots e^{t_{nn}} = e^{\operatorname{tr} T} = e^{\operatorname{tr} X},$$

since $\operatorname{tr} T = \operatorname{tr}(Y(XY^{-1})) = \operatorname{tr}((XY^{-1})Y) = \operatorname{tr} X$. □

Exercise 3.29. Prove Proposition 3.28 (i)–(v).

Definition 3.30. If the $n \times n$ matrices X and Y over $\mathbb{R}$ or $\mathbb{C}$ satisfy $\exp Y = X$, then $Y = \log X$ is called a *matrix logarithm* of X.

$\log X$ does not exist for every matrix X, and if it exists it is usually not unique. For $X = (x_{ij})$,

$$\|X\|_{\mathrm{Frob}} := \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} |x_{ij}|^2}$$

is called the *Frobenius norm* or *Schur norm* of X. If $\|X\|_{\mathrm{Frob}} < 1$, then

$$\log(I - X) = -\sum_{k=1}^{\infty} \frac{1}{k} X^k \tag{4.8}$$

is a matrix logarithm of X.

Theorem 3.31 (Two-term determinant formula). *Let $\mathcal{G}$ be a friendly graph and let $\vec{\mathcal{G}}$ be an orientation of $\mathcal{G}$ with edge adjacency matrix $W_{\vec{\mathcal{G}}}$. Then*

$$\zeta_{\mathcal{G}}(u)^{-1} = \det(I - W_{\vec{\mathcal{G}}}u).$$

Proof. Let $\mathcal{G} = (V, E)$ and $m := |E|$. Recall that $N_m(\mathcal{G})$ is the number of backtrackless cycles of length m in $\mathcal{G}$. We first claim that

$$N_m(\mathcal{G}) = \operatorname{tr} W_{\vec{\mathcal{G}}}^m. \tag{4.9}$$

To prove this, let $W_{\vec{\mathcal{G}}} = (w_{ij})$ and $W_{\vec{\mathcal{G}}}^k = (w_{ij}^{(k)})$ for all positive k. By induction,

$$w_{ij}^{(m)} = \sum_{\ell_1}\sum_{\ell_2}\cdots\sum_{\ell_{m-1}} w_{i\ell_1}w_{\ell_1\ell_2}\ldots w_{\ell_{m-1}j},$$

hence

$$\operatorname{tr} W_{\vec{\mathcal{G}}}^m = \sum_i w_{ii}^{(m)} = \sum_i\sum_{\ell_1}\sum_{\ell_2}\cdots\sum_{\ell_{m-1}} w_{i\ell_1}w_{\ell_1\ell_2}\ldots w_{\ell_{m-1}i}.$$

Since we have $w_{i\ell_1}w_{\ell_1\ell_2}\ldots w_{\ell_{m-1}i} = 1$ if and only if $e_i e_{\ell_1} e_{\ell_2}\ldots e_{\ell_{m-1}}$ is a closed path of length m starting with edge e_i, $\operatorname{tr} W_{\vec{\mathcal{G}}}^m$ apparently counts all these paths, and just the ones without backtracks due to the definition of w_{ij}. This proves (4.9).

By virtue of identities (4.4) in Proposition 3.11 and (4.9), we have for sufficiently small $|u|$

$$\log \zeta_{\mathcal{G}}(u) = \sum_{m=1}^{\infty} \frac{N_m(\mathcal{G})}{m} u^m = \sum_{m=1}^{\infty} \frac{u^m}{m} \operatorname{tr} W_{\vec{\mathcal{G}}}^m = \operatorname{tr}\left(\sum_{m=1}^{\infty} \frac{1}{m}(W_{\vec{\mathcal{G}}}u)^m\right), \tag{4.10}$$

where the last identity follows from the fact that tr is continous and linear.

If $|u|$ is sufficiently small, we clearly have $\|W_{\vec{\mathcal{G}}}u\|_{\mathrm{Frob}} < 1$, and consequently

$$\sum_{m=1}^{\infty} \frac{1}{m}\,(W_{\vec{\mathcal{G}}}u)^m = -\log(I - W_{\vec{\mathcal{G}}}u)$$

by (4.8). This and (4.10) imply

$$\zeta_{\mathcal{G}}(u)^{-1} = e^{\operatorname{tr}\left(\log(I-W_{\vec{\mathcal{G}}}u)\right)} = \det\left(\exp\left(\log(I - W_{\vec{\mathcal{G}}}u)\right)\right) = \det(I - W_{\vec{\mathcal{G}}}u), \quad (4.11)$$

thanks to Proposition 3.28 (vi). □

Corollary 3.32. *For a friendly graph $\mathcal{G}$ with orientation $\vec{\mathcal{G}}$ and edge adjacency matrix $W_{\vec{\mathcal{G}}}$, we have*

$$\zeta_{\mathcal{G}}(u)^{-1} = \prod_{\lambda\in\mathrm{Spec}(W_{\vec{\mathcal{G}}})} (1 - \lambda u).$$

In particular, the poles of $\zeta_{\mathcal{G}}(u)$ are the reciprocals of the eigenvalues of $W_{\vec{\mathcal{G}}}$.

Proof. $\det(zI - W_{\vec{\mathcal{G}}})$ is the monic characteristic polynomial of $W_{\vec{\mathcal{G}}}$, i.e.,

$$\det(zI - W_{\vec{\mathcal{G}}}) = \prod_{\lambda\in\mathrm{Spec}(W_{\vec{\mathcal{G}}})} (z - \lambda).$$

For the number m of edges in $\mathcal{G}$, we consequently have

$$\begin{aligned}\det(I - W_{\vec{\mathcal{G}}}u) = \det\left(u\left(\frac{1}{u}I - W_{\vec{\mathcal{G}}}\right)\right) &= u^{2m} \prod_{\lambda\in\mathrm{Spec}(W_{\vec{\mathcal{G}}})} \left(\frac{1}{u} - \lambda\right)\\ &= \prod_{\lambda\in\mathrm{Spec}(W_{\vec{\mathcal{G}}})} (1 - \lambda u).\end{aligned}$$ □

The last ingredients we need to verify Theorem 3.15 are a few identities for the following block matrices related to an orientation $\vec{\mathcal{G}}$ of a graph $\mathcal{G} = (V, E)$ with $n := |V|$ and $m := |E|$:

- The $2m \times 2m$ matrix
$$J := \begin{pmatrix} O_m & I_m \\ I_m & O_m \end{pmatrix},$$
where O_m denotes the $m \times m$ zero matrix and I_m is the $m \times m$ unit matrix;
- the $n \times 2m$ *start matrix* $S_{\vec{\mathcal{G}}} = (s_{ij})$, derived by
$$s_{ij} := \begin{cases} 1, & \text{if } v_i \text{ is the starting vertex of } e_j \text{ in } \vec{\mathcal{G}},\\ 0, & \text{otherwise;}\end{cases}$$

- the $n \times 2m$ *terminal matrix* $T_{\vec{\mathcal{G}}} = (t_{ij})$ defined by

$$t_{ij} := \begin{cases} 1, & \text{if } v_i \text{ is the terminal vertex of } e_j \text{ in } \vec{\mathcal{G}}, \\ 0, & \text{otherwise.} \end{cases}$$

Note that, upon numbering directed edges in our orientation $\vec{\mathcal{G}}$ of $\mathcal{G}$ by the rule $e_{j+m} := e_j^{-1}$, we have $S = (M|N)$ and $T = (N|M)$ for two suitable $n \times m$ matrices M and N.

Proposition 3.33. *Let $\mathcal{G} = (V, E)$ be a graph with $n := |V|$ and $m := |E|$, and let $\vec{\mathcal{G}}$ be an orientation of $\mathcal{G}$. Let J, $S_{\vec{\mathcal{G}}}$, and $T_{\vec{\mathcal{G}}}$ be defined as above, $A_{\mathcal{G}}$ be the adjacency matrix of $\mathcal{G}$, $W_{\vec{\mathcal{G}}}$ be the edge adjacency matrix of $\vec{\mathcal{G}}$, and $Q_{\mathcal{G}}$ be as defined in Theorem 3.15. Then*

(i) $S_{\vec{\mathcal{G}}} J = T_{\vec{\mathcal{G}}}$ *and* $T_{\vec{\mathcal{G}}} J = S_{\vec{\mathcal{G}}}$;

(ii) $A_{\mathcal{G}} = S_{\vec{\mathcal{G}}} T_{\vec{\mathcal{G}}}^t$ *and* $Q_{\mathcal{G}} + I_n = S_{\vec{\mathcal{G}}} S_{\vec{\mathcal{G}}}^t = T_{\vec{\mathcal{G}}} T_{\vec{\mathcal{G}}}^t$;

(iii) $W_{\vec{\mathcal{G}}} + J = T_{\vec{\mathcal{G}}}^t S_{\vec{\mathcal{G}}}$.

Partial proof. The proofs of (i) and the second identity in (ii) are left as exercises. By the definition of matrix multiplication,

$$\left(S_{\vec{\mathcal{G}}} T_{\vec{\mathcal{G}}}^t\right)_{ik} = \sum_{j=1}^{2m} s_{ij} t_{kj}, \tag{4.12}$$

where $s_{ij} t_{kj} = 1$ if and only if the edge e_j has starting vertex v_i and terminal vertex v_k, and $s_{ij} t_{kj} = 0$ otherwise. Hence the right-hand side of (4.12) equals the number of oriented edges from v_i to v_k, and precisely this is the entry A_{ij}. This proves the first identity in (ii).

Similarly, we have

$$\left(T_{\vec{\mathcal{G}}}^t S_{\vec{\mathcal{G}}}\right)_{ik} = \sum_{j=1}^{n} t_{ji} s_{jk},$$

where $t_{ji} s_{jk} = 1$ if and only if the edge e_i feeds via the vertex v_j directly into the edge e_k (and this is even true in case $e_k = e_i^{-1}$), and $t_{ji} s_{jk} = 0$ otherwise. Hence

$$\sum_{j=1}^{n} t_{ji} s_{jk} = \begin{cases} 1, & \text{if } e_i \text{ feeds directly into } e_k, \\ 0, & \text{otherwise.} \end{cases} \tag{4.13}$$

By definition, this equals the entry w_{ik} of $W_{\vec{\mathcal{G}}}$ for $e_k \neq e_i^{-1}$. Recalling our labeling convention $e_{j+m} := e_j^{-1}$ in $\vec{\mathcal{G}}$, we see that the right-hand side of (4.13) equals $w_{ik} + 1$ in case $e_k = e_i^{-1}$. Putting everything together, (iii) follows. □

Exercise 3.34.

(i) Prove (i) and the relation $Q_{\mathcal{G}} + I_n = S_{\vec{\mathcal{G}}} S_{\vec{\mathcal{G}}}^t = T_{\vec{\mathcal{G}}} T_{\vec{\mathcal{G}}}^t$ from Proposition 3.33.

(ii) Show that $A_{\mathcal{G}} = S_{\vec{\mathcal{G}}} T_{\vec{\mathcal{G}}}^t$ even holds if $\mathcal{G}$ is allowed to have loops.

(iii) Use Prop. 3.33 to create random graphs (see figure at the end of Section 3.2): Take random permutation matrices $P_1, \ldots, P_k$ and define $M := P_1 \cdot \cdots \cdot P_k$. Similarly, build N, and then define $S_{\vec{\mathcal{G}}} := (M|N)$ and $T_{\vec{\mathcal{G}}} := (N|M)$. Finally, obtain $W_{\vec{\mathcal{G}}}$ by Proposition 3.33 (iii).

Proof of the three-term determinant formula – Theorem 3.15. Let $n = |V|$ and $m = |E|$, and let $A := A_{\mathcal{G}}$, $Q := Q_{\mathcal{G}}$, $W := W_{\vec{\mathcal{G}}}$, $S := S_{\vec{\mathcal{G}}}$ and $T := T_{\vec{\mathcal{G}}}$. All four-block matrices in the following calculations are of size $(n + 2m) \times (n + 2m)$ with blocks of size $n \times n$ in the upper left corner. Then

$$\begin{aligned} M_1 &:= \begin{pmatrix} I_n & O \\ T^t & I_{2m} \end{pmatrix} \cdot \begin{pmatrix} I_n(1-u^2) & Su \\ O & I_{2m} - Wu \end{pmatrix} \\ &= \begin{pmatrix} I_n(1-u^2) & Su \\ T^t(1-u^2) & T^tSu + I_{2m} - Wu \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} M_2 &:= \begin{pmatrix} I_n - Au + Qu^2 & Su \\ O & I_{2m} + Ju \end{pmatrix} \cdot \begin{pmatrix} I_n & O \\ T^t - S^tu & I_{2m} \end{pmatrix} \\ &= \begin{pmatrix} I_n - Au + Qu^2 + ST^tu - SS^tu^2 & Su \\ (I_{2m} + Ju)(T^t - S^tu) & I_{2m} + Ju \end{pmatrix}. \end{aligned}$$

We claim that $M_1 = M_2$. For the upper left corner of M_2 we have, by Proposition 3.33,

$$\begin{aligned} I_n - Au + Qu^2 + ST^tu - SS^tu^2 &\overset{\text{(ii)}}{=} I_n - ST^tu + (SS^t - I_n)u^2 + ST^tu - SS^tu^2 \\ &= I_n(1-u^2). \end{aligned}$$

Similarly, the lower left corner of M_2 equals

$$\begin{aligned} (I_{2m} + Ju)(T^t - S^tu) &= T^t - S^tu + JT^tu - JS^tu^2 \\ &= T^t - S^tu + (TJ)^tu - (SJ)^tu^2 \\ &\overset{\text{(i)}}{=} T^t - S^tu + S^tu - T^tu^2 = \\ &= T^t(1-u^2), \end{aligned}$$

and finally the lower right corner of M_1 is

$$\begin{aligned} T^tSu + I_{2m} - Wu &\overset{\text{(iii)}}{=} T^tSu + I_{2m} - (T^tS - J)u \\ &= I_{2m} + Ju, \end{aligned}$$

which proves our claim. Consequently, $\det M_1 = \det M_2$ with

$$\begin{aligned}\det M_1 &= \det(I_n(1-u^2))\det(I_{2m} - Wu)\\ &= (1-u^2)^n \det(I_{2m} - Wu) = (1-u^2)^n \zeta_{\mathcal{G}}(u)^{-1}\end{aligned}$$

by the two-determinant formula Theorem 3.31. On the other hand,

$$\det M_2 = \det(I_n - Au + Qu^2)\det(I_{2m} + Ju),$$

which implies

$$\begin{aligned}\zeta_{\mathcal{G}}(u)^{-1} &= (1-u^2)^{-n}\det M_1 = (1-u^2)^{-n}\det M_2\\ &= (1-u^2)^{-n}\det(I_n - Au + Qu^2)\det(I_{2m} + Ju).\end{aligned} \tag{4.14}$$

Observe that

$$\begin{pmatrix} I_m & O \\ -I_m u & I_m\end{pmatrix}\cdot(I_{2m}+Ju) = \begin{pmatrix} I_m & O \\ -I_m u & I_m\end{pmatrix}\cdot\begin{pmatrix} I_m & I_m u \\ I_m u & I_m\end{pmatrix} = \begin{pmatrix} I_m & I_m u \\ O & I_m(1-u^2)\end{pmatrix},$$

hence

$$\det(I_{2m} + Ju) = \det\begin{pmatrix} I_m & I_m u \\ O & I_m(1-u^2)\end{pmatrix} = (1-u^2)^m.$$

By (4.14), this yields

$$\zeta_{\mathcal{G}}(u)^{-1} = (1-u^2)^{m-n}\det(I_n - Au + Qu^2) = (1-u^2)^{r-1}\det(I_n - Au + Qu^2)$$

for connected graphs, by Proposition 3.13 (ii). □

3.4 The prime number theorem for graphs For a friendly graph $\mathcal{G}$, we defined $\mathbb{P}_{\mathcal{G}}$ as the set of primes in $\mathcal{G}$, the greatest common divisor $\Delta_{\mathcal{G}} = \gcd\{\nu[P] : [P] \in \mathbb{P}_{\mathcal{G}}\}$ and the prime counting function

$$\pi_{\mathcal{G}}(k) := \#\{[P] \in \mathbb{P}_{\mathcal{G}} : \ \nu[P] = k\}.$$

Recall that $R_{\mathcal{G}}$ denotes the radius of convergence of $\zeta_{\mathcal{G}}$ and is the reciprocal of the modulus of the largest eigenvalue of $W_{\vec{\mathcal{G}}}$ (cf. Corollary 3.32), while the second largest modulus of the eigenvalues is $\Lambda(W_{\vec{\mathcal{G}}}) := \max\{|\ \lambda\ |: \ \lambda \in \operatorname{Spec}(W_{\vec{\mathcal{G}}}),\ |\ \lambda\ |< R_{\mathcal{G}}^{-1}\}$ (cf. Definiton 2.34).

Theorem 3.35. *Let $\mathcal{G} = (V, E)$ be a friendly graph, and let k be a positive integer. If $\Delta_{\mathcal{G}} \nmid k$, then $\pi_{\mathcal{G}}(k) = 0$. For $\Delta_{\mathcal{G}} \mid k$, we have*

$$\pi_{\mathcal{G}}(k) = \Delta_{\mathcal{G}}\frac{R_{\mathcal{G}}^{-k}}{k} + O\Big(\frac{\Lambda(W_{\vec{\mathcal{G}}})^k}{k}\Big) + O\Big(\frac{R_{\mathcal{G}}^{-\frac{k}{2}}}{k}\Big). \tag{4.15}$$

In particular, we have $\pi_{\mathcal{G}}(k) \sim \Delta_{\mathcal{G}}\frac{R_{\mathcal{G}}^{-k}}{k}$ asymptotically for $\Delta_{\mathcal{G}} \mid k$ with $k \to \infty$.

Proof. This proof somewhat imitates the proof of the analogous result for zeta functions of function fields introduced by Emil Artin — where the Riemann hypothesis is known to be true by works of Helmut Hasse, André Weil and Pierre Deligne (cf. [38]).

By Proposition 3.8, the case $\Delta_{\mathcal{G}} \nmid k$ has already been treated, and we may assume that $\Delta_{\mathcal{G}} \mid k$. By our definitions,

$$\zeta_{\mathcal{G}}(u) = \prod_{[P]\in\mathbb{P}_{\mathcal{G}}} \left(1 - u^{\nu[P]}\right)^{-1} = \prod_{k=1}^{\infty} \left(1 - u^{k}\right)^{-\pi_{\mathcal{G}}(k)}.$$

This implies that

$$\begin{aligned}\log \zeta_{\mathcal{G}}(u) &= -\sum_{k=1}^{\infty} \pi_{\mathcal{G}}(k) \log(1-u^{k})\\ &= \sum_{k=1}^{\infty} \pi_{\mathcal{G}}(k) \sum_{\ell=1}^{\infty} \frac{u^{k\ell}}{\ell} \overset{k\ell=m}{=} \sum_{m=1}^{\infty} u^{m} \sum_{k\mid m} \pi_{\mathcal{G}}(k) \cdot \frac{k}{m},\end{aligned}$$

and consequently

$$u \frac{d}{du} \log \zeta_{\mathcal{G}}(u) = u \sum_{m=1}^{\infty} m u^{m-1} \sum_{k\mid m} \pi_{\mathcal{G}}(k) \cdot \frac{k}{m} = \sum_{m=1}^{\infty} u^{m} \sum_{k\mid m} k\pi_{\mathcal{G}}(k). \qquad (4.16)$$

On the other hand, we know from (4.4) in Proposition 3.11 that

$$u \frac{d}{du} \log \zeta_{\mathcal{G}}(u) = u \frac{d}{du}\Big(\sum_{m=1}^{\infty} \frac{N_m(\mathcal{G})}{m} u^{m}\Big) = \sum_{m=1}^{\infty} N_m(\mathcal{G}) u^{m}, \qquad (4.17)$$

which combined with (4.16) yields

$$N_m(\mathcal{G}) = \sum_{k\mid m} k\pi_{\mathcal{G}}(k). \qquad (4.18)$$

Using Möbius' inversion formula, we obtain that

$$\pi_{\mathcal{G}}(k) = \frac{1}{k} \sum_{d\mid k} \mu\Big(\frac{k}{d}\Big) N_d(\mathcal{G}) = \frac{1}{k}\Big(N_k(\mathcal{G}) + \sum_{d\mid k,\, d\le \frac{k}{2}} \mu\Big(\frac{k}{d}\Big) N_d(\mathcal{G})\Big). \qquad (4.19)$$

Applying the two-term determinant formula in terms of Corollary 3.32, we deduce a third identity for $u \frac{d}{du} \log \zeta_{\mathcal{G}}(u)$, namely

$$\begin{aligned}u \frac{d}{du} \log \zeta_{\mathcal{G}}(u) &= u \frac{d}{du} \log \Big(\prod_{\lambda\in\mathrm{Spec}(W_{\vec{\mathcal{G}}})} (1-\lambda u)\Big)^{-1}\\ &= -u \frac{d}{du} \sum_{\lambda\in\mathrm{Spec}(W_{\vec{\mathcal{G}}})} \log(1-\lambda u)\end{aligned}$$

$$= u\,\frac{d}{du} \sum_{\lambda \in \mathrm{Spec}(W_{\vec{\mathcal{G}}})} \sum_{m=1}^{\infty} \frac{(\lambda u)^m}{m}$$

$$= \sum_{\lambda \in \mathrm{Spec}(W_{\vec{\mathcal{G}}})} \sum_{m=1}^{\infty} (\lambda u)^m$$

$$= \sum_{m=1}^{\infty} \Big(\sum_{\lambda \in \mathrm{Spec}(W_{\vec{\mathcal{G}}})} \lambda^m \Big) u^m .$$

By comparing with (4.17), we see that

$$N_m(\mathcal{G}) = \sum_{\lambda \in \mathrm{Spec}(W_{\vec{\mathcal{G}}})} \lambda^m . \tag{4.20}$$

For large m the dominant terms in the last sum are the ones coming from $\lambda \in \mathrm{Spec}(W_{\vec{\mathcal{G}}})$ with maximal absolute value, i.e., from the poles u of $\zeta_{\mathcal{G}}(u)$ with minimal absolute value, as follows from Corollary 3.32. By Theorem 3.24 (i), these λ have absolute value $|\lambda| = R_{\mathcal{G}}^{-1}$. By Theorem 3.24 (iii), we even know them exactly and they have the form

$$\lambda = R_{\mathcal{G}}^{-1} e^{\frac{2\pi i k}{\Delta_{\mathcal{G}}}} \qquad (k = 1, \ldots, \Delta_{\mathcal{G}}).$$

Hence, by (4.20),

$$N_m(\mathcal{G}) = \sum_{k=1}^{\Delta_{\mathcal{G}}} R_{\mathcal{G}}^{-m} e^{\frac{2\pi i k m}{\Delta_{\mathcal{G}}}} + \sum_{\lambda \in \mathrm{Spec}(W_{\vec{\mathcal{G}}}),\, |\lambda| \le \Lambda(W_{\vec{\mathcal{G}}})} \lambda^m . \tag{4.21}$$

For the main term in (4.21) we obtain

$$\sum_{k=1}^{\Delta_{\mathcal{G}}} R_{\mathcal{G}}^{-m} e^{\frac{2\pi i k m}{\Delta_{\mathcal{G}}}} = R_{\mathcal{G}}^{-m} \sum_{k=1}^{\Delta_{\mathcal{G}}} e^{\frac{2\pi i k m}{\Delta_{\mathcal{G}}}} = \begin{cases} R_{\mathcal{G}}^{-m} \Delta_{\mathcal{G}}, & \text{if } \Delta_{\mathcal{G}} \mid m, \\ 0, & \text{if } \Delta_{\mathcal{G}} \nmid m, \end{cases} \tag{4.22}$$

and the error term can be bounded as

$$\Bigg| \sum_{\lambda \in \mathrm{Spec}(W_{\vec{\mathcal{G}}}),\, |\lambda| \le \Lambda(W_{\vec{\mathcal{G}}})} \lambda^m \Bigg| \le |\mathrm{Spec}(W_{\vec{\mathcal{G}}})| \Lambda(\mathcal{G})^m = 2|E| \Lambda(W_{\vec{\mathcal{G}}})^m . \tag{4.23}$$

Combining (4.21), (4.22), and (4.23), we obtain in case $\Delta_{\mathcal{G}} \mid k$, as we may assume, that

$$N_k(\mathcal{G}) = R_{\mathcal{G}}^{-k} \Delta_{\mathcal{G}} + 2|E| \Lambda(W_{\vec{\mathcal{G}}})^k = R_{\mathcal{G}}^{-k} \Delta_{\mathcal{G}} + O(\Lambda(W_{\vec{\mathcal{G}}})^k), \tag{4.24}$$

and for all d

$$|N_d(\mathcal{G})| \leq R_{\mathcal{G}}^{-d}\Delta_{\mathcal{G}} + 2|E|\Lambda(W_{\vec{\mathcal{G}}})^d = O(R_{\mathcal{G}}^{-d}), \tag{4.25}$$

since $|R_{\mathcal{G}}^{-1}| > \Lambda(W_{\vec{\mathcal{G}}})$ by definition. Inserting (4.24) and (4.25) into (4.19) yields

$$\begin{aligned}\pi_{\mathcal{G}}(k) &= \frac{1}{k}\Big(R_{\mathcal{G}}^{-k}\Delta_{\mathcal{G}} + O(\Lambda(W_{\vec{\mathcal{G}}})^k) + O\Big(\sum_{1\leq d\leq \frac{k}{2}} R_{\mathcal{G}}^{-d}\Big)\Big)\\ &= \Delta_{\mathcal{G}}\frac{R_{\mathcal{G}}^{-k}}{k} + O\Big(\frac{\Lambda(W_{\vec{\mathcal{G}}})^k}{k}\Big) + O\Big(\frac{R_{\mathcal{G}}^{-\frac{k}{2}}}{k}\Big). \qquad \square\end{aligned}$$

Example 3.36. Reconsider the graph $\mathcal{K}_4 - e$ (cf. Example 3.2 and Exercise 3.17) with

$$\zeta_{\mathcal{K}_4-e}(u)^{-1} = (1-u^2)(1-u)(1+u^2)(1+u+2u^2)(1-u^2-2u^3).$$

Using (4.17), we obtain after a little computation

$$\begin{aligned}\sum_{m=1}^{\infty} N_m(\mathcal{G})u^m &= u\,\frac{d}{du}\log\zeta_{\mathcal{G}}(u)\\ &= 12u^3 + 8u^4 + 24u^6 + 28u^7 + 8u^8 + 48u^9 + 120u^{10} +\\ &\quad + 44u^{11} + 104u^{12} + 416u^{13} + 280u^{14} + O(u^{15}).\end{aligned}$$

Now we apply the formula $N_m = N_m(\mathcal{K}_4 - e) = \sum_{k|m} k\pi_{\mathcal{K}_4-e}(k)$ in (4.18) to compute small values of $\pi(k) := \pi_{\mathcal{K}_4-e}(k)$:

$$\begin{array}{ll} 12 = N_3 = 1\cdot\pi(1) + 3\cdot\pi(3) = 3\pi(3) & \Longrightarrow \pi(3) = 4;\\ 8 = N_4 = \pi(1) + 2\pi(2) + 4\pi(4) = 4\pi(4) & \Longrightarrow \pi(4) = 2;\\ 0 = N_5 = \pi(1) + 5\pi(5) = 5\pi(5) & \Longrightarrow \pi(5) = 0;\\ 24 = N_6 = \pi(1) + 2\pi(2) + 3\pi(3) + 6\pi(6) = 12 + 6\pi(6) & \Longrightarrow \pi(6) = 2;\\ 28 = N_7 = \pi(1) + 7\pi(7) = 7\pi(7) & \Longrightarrow \pi(7) = 4;\\ 8 = N_8 = \pi(1) + 2\pi(2) = 4\pi(4) + 8\pi(8) = 8 + 8\pi(8) & \Longrightarrow \pi(8) = 0;\\ 48 = N_9 = \pi(1) + 3\pi(3) + 9\pi(9) = 12 + 9\pi(9) & \Longrightarrow \pi(9) = 4;\\ 120 = N_{10} = \pi(1) + 2\pi(2) + 5\pi(5) + 10\pi(10) = 10\pi(10) & \Longrightarrow \pi(10) = 12.\end{array}$$

Exercise 3.37.

(i) Identify the different primes in $\mathcal{K}_4 - e$ according to Example 3.36.

(ii) Use Example 3.16 to find $\pi_{\mathcal{K}_4}(k)$ for $k = 3, 4, 5, \ldots, 11$.

3.5 Some remarks on Ramanujan graphs At the end of Section 2.2 we introduced Ramanujan graphs as special expander graphs having applications in communication network theory (cf. [6]), and theoretical computer science (cf. [19]) in general. In this section we emphasize their importance regarding the Riemann hypothesis for graphs, i.e., the distribution of primes in graphs.

The asymptotic formula (4.15) has a "good" error term if the second largest eigenvalue of the edge adjacency matrix of the graph $\mathcal{G}$ is "small". For $(q+1)$-regular graphs, in particular, this is true when the graph is a Ramanujan graph with $\Lambda(\mathcal{G}) \leq 2\sqrt{q}$, and this in turn is equivalent with the fact that the corresponding Riemann hypothesis is satisfied (see Theorem 3.20). The following theorem in fact shows that in this case the error term will be best possible if we consider a family of $(q+1)$-regular graphs whose vertex numbers go to infinity.

Theorem 3.38 (Alon [2] and Boppana [8], 1986–87). *Let $(\mathcal{G}_n)_{n\in\mathbb{N}}$ be a sequence of $(q+1)$-regular graphs $\mathcal{G}_n = (V_n, E_n)$ with $\lim_{n\to\infty} |V_n| = \infty$. Then*

$$\lim_{n\to\infty} \inf_{m\geq n} \Lambda(\mathcal{G}_m) \geq 2\sqrt{q}.$$

In 1988 Lubotzky, Phillips, and Sarnak [31] constructed such an infinite family of $(q+1)$-regular Ramanujan graphs for $q \equiv 1 \bmod 4$ prime. Their proof uses the Ramanujan conjecture, which led to the name of Ramanujan graphs. Morgenstern [34] extended the construction of Lubotzky, Phillips, and Sarnak to all prime powers in 1994. The problem is open for general q.

In Section 4.2 we shall take up the topic of Ramanujan graphs once again.

4 Spectral properties of integral circulant graphs

A graph is called *integral* if its spectrum is integral (cf. Definition 2.32). Up to now no handy criterion for this property is known (cf. [16]). Yet many classes of integral graphs have been identified and studied recently, in particular in the case when the graph is *circulant* (cf. Definition 2.28), meaning its adjacency matrix is circulant (see e.g. [1], [3], [22], [42], [45]). The class of *integral circulant graphs* displays rich algebraic, arithmetic and combinatorial features. In this section we shall focus on some arithmetical aspects of spectra of these graphs, including open problems and conjectures. Our main topics will be the Ramanujan property (see Definition 2.34) discussed in Section 3.5, and the concept of the *energy* of a graph.

4.1 Basics on integral circulant graphs By Theorem 2.33 we know that integral circulant graphs $\mathrm{ICG}(n, \mathcal{D}) = (\mathbb{Z}_n, E)$ are Cayley graphs characterized by their order n and a set $\mathcal{D}$ of positive divisors of n and

$$E := \{(a,b) : a, b \in \mathbb{Z}_n, (a-b, n) \in \mathcal{D}\}.$$

The unitary Cayley graphs (cf. Definition 2.13 (ii)) are the integral circulant graphs with $\mathcal{D} = \{1\}$.

Let us recall some general facts about $\mathrm{ICG}(n, \mathcal{D})$ for arbitrary positive integers n and arbitrary divisor sets $\mathcal{D} \subseteq D(n) := \{d > 0 : d \mid n\}$.

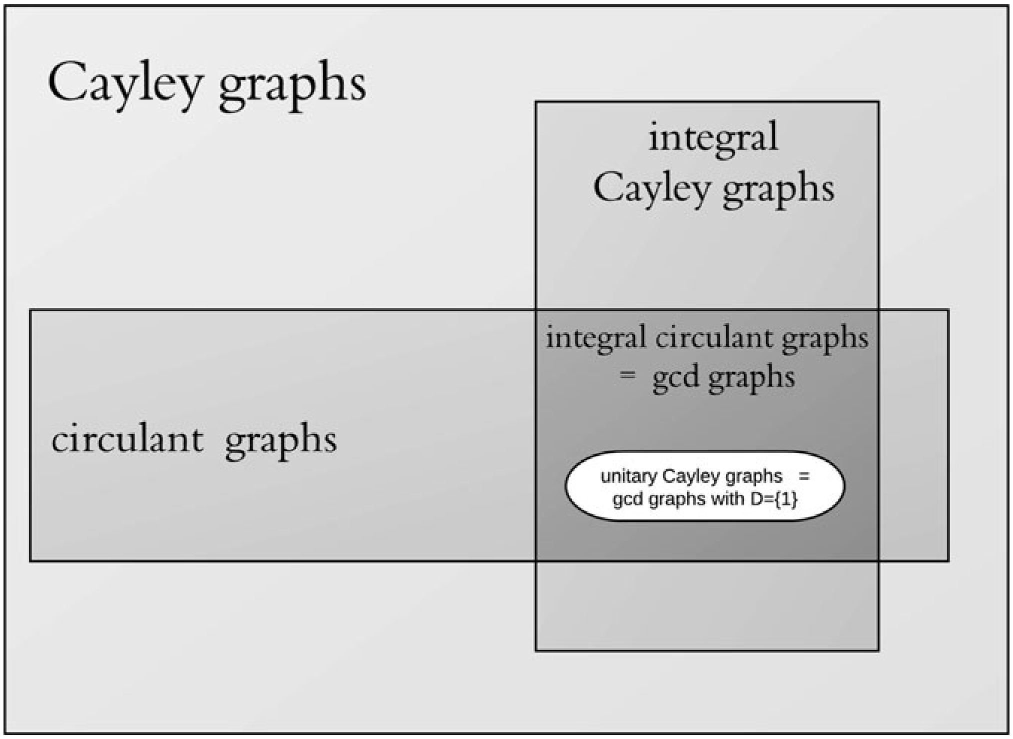


Figure 4.15. Subclasses of graphs among Cayley graphs

Proposition 4.1.

(i) *Since* ICG$(n, \mathcal{D})$ *is circulant, its eigenvalues* $\lambda_k(n, \mathcal{D})$ $(1 \le k \le n)$*, say, can be calculated by using Proposition* 2.31*, yielding*

$$\lambda_k(n, \mathcal{D}) = \sum_{d \in \mathcal{D}} c\Big(k, \frac{n}{d}\Big) \qquad (1 \le k \le n), \tag{4.26}$$

where

$$c(k, n) := \sum_{\substack{j \bmod n \\ (j,n)=1}} \exp\Big(\frac{2\pi i\, kj}{n}\Big)$$

is the well-known ***Ramanujan sum*** (*cf.* [4] *or* [44]).

(ii) ICG$(n, \mathcal{D})$ *is regular, more precisely* $\Phi(n, \mathcal{D})$*-regular, where*

$$\Phi(n, \mathcal{D}) := \lambda_n(n, \mathcal{D}) = \sum_{d \in \mathcal{D}} \varphi\Big(\frac{n}{d}\Big), \tag{4.27}$$

with Euler's totient function φ*. This follows from Exercise* 2.15 (iii)*, saying that a Cayley graph* $\mathrm{Cay}(G, S)$ *is* $|S|$*-regular.*

(iii) *By the observation of So* [46] *that* $\mathrm{ICG}(n, \mathcal{D})$ *with* $\mathcal{D} = \{d_1, \ldots, d_r\}$*, say, is connected if and only if* $\gcd(d_1, \ldots, d_r) = 1$*, connectivity can readily be checked. If* $\mathrm{ICG}(n, \mathcal{D})$ *is connected, then* $\Phi(n, \mathcal{D})$ *is the largest eigenvalue of* $\mathrm{ICG}(n, \mathcal{D})$*, the so-called spectral radius of the graph, and it occurs with multiplicity* 1 (*cf. Proposition* 2.26 (ii)).

Exercise 4.2.

(i) Show that $\mathrm{ICG}(n, \mathcal{D})$ has loops if and only if $n \in \mathcal{D}$. For that reason, we shall usually require that $\mathcal{D} \subseteq D^*(n) := \{0 < d < n : \ d \mid n\}$.

(ii) Deduce (4.26) from Proposition 2.31.

(iii) Prove Proposition 4.1 (ii).

Recently, Le and the author [28] observed that (4.26) can be rewritten as

$$\lambda_k(n, \mathcal{D}) = (\mathbb{1} *_{\mathcal{D}} c(k, \cdot))(n) \qquad (1 \le k \le n), \tag{4.28}$$

where $\mathbb{1}$ is the constant function and $*_{\mathcal{D}}$ denotes the so-called $\mathcal{D}$-*convolution* of arithmetic functions introduced by Narkiewicz [36], which is a generalisation of the classical Dirichlet convolution. This representation of the eigenvalues will be of great importance for our purpose.

In general, given non-empty sets $A(n) \subseteq D(n)$ for all positive integers n, the (*arithmetical*) *convolution* A or the A-*convolution* of two arithmetic functions $f, g \in \mathbb{C}^{\mathbb{N}}$ is defined as

$$(f *_A g)(n) = \sum_{d \in A(n)} f(d) g\Big(\frac{n}{d}\Big).$$

All convolutions considered in the literature are required to be *regular*, a property which basically guarantees that $f *_A g$ is multiplicative for multiplicative functions f and g, and the inverse of $\mathbb{1}$ with respect to $*_A$, i.e., an analogue of the Möbius function μ, exists. Narkiewicz [36] proved that regularity of an A-convolution is, besides some minor technical requirements, essentially equivalent with the following two conditions:

(i) A is *multiplicative*, i.e., $A(mn) = A(m)A(n) := \{ab : \ a \in A(m), b \in A(n)\}$ for all coprime $m, n \in \mathbb{N}$.

(ii) A is *semi-regular*, i.e., for every prime power p^s with $s \ge 1$ there exists a divisor $t = t_A(p^s)$ of s, called the *type* of p^s, such that $A(p^s) = \{1, p^t, p^{2t}, \ldots, p^{jt}\}$ with $j = \frac{s}{t}$.

Both of these properties will occur in a natural fashion along our way, but for our purposes concerning the eigenvalues of $\mathrm{ICG}(n, \mathcal{D})$ the multiplicativity of the divisor sets $\mathcal{D}$ will be the guiding feature.

The product of non-empty sets $A_1, \ldots, A_t$ of integers is defined as

$$\prod_{i=1}^{t} A_i := \{a_1 \cdot \cdots \cdot a_t : \ a_i \in A_i \ \ (1 \le i \le t)\}.$$

For infinitely many such sets $A_1, A_2, \ldots$, we require that $A_i = \{1\}$ for all but finitely many i and define

$$\prod_{i=1}^{\infty} A_i := \prod_{\substack{i=1 \\ A_i \neq \{1\}}}^{\infty} A_i$$

Let us call a set $\mathcal{A}$ of positive integers a *multiplicative set* if it is the product of non-empty finite sets $A_i \subset \{1, p_i, p_i^2, p_i^3, \ldots\}$, $1 \leq i \leq t$ say, with pairwise distinct primes $p_1, \ldots, p_t$. In other words, a given set $\mathcal{A}$ is multiplicative if and only if $\mathcal{A} = \prod_{p \in \mathbb{P}} \mathcal{A}_p$, where $\mathcal{A}_p := \{p^{e_p(a)} : a \in \mathcal{A}\}$ for each prime p, and $e_p(a)$ denotes the order of the prime p in a. Observe that $\mathcal{A}_p \neq \{1\}$ only for those finitely many primes dividing at least one of the $a \in \mathcal{A}$.

If $\mathcal{D}$ is a multiplicative divisor set, then this property is extended to the spectrum of the corresponding integral circulant graph.

Proposition 4.3 (Le and Sander [28], 2012). *For a multiplicative divisor set $\mathcal{D} = \prod_{p|n} \mathcal{D}_p \subseteq D(n)$,*

$$\lambda_k(n, \mathcal{D}) = \prod_{p \in \mathbb{P},\, p|n} \lambda_k(p^{e_p(n)}, \mathcal{D}_p), \tag{4.29}$$

where for each prime p dividing n the integers $\lambda_k(p^{e_p(n)}, \mathcal{D}_p)$, $1 \leq k \leq p^{e_p(n)}$, are the eigenvalues of ICG$(p^{e_p(n)}, \mathcal{D}_p)$*, and $\lambda_k(p^{e_p(n)}, \mathcal{D}_p)$ is defined for all integers k by periodic continuation* mod $p^{e_p(n)}$.

4.2 Ramanujan integral circulant graphs

In 2010 Droll [13] classified all Ramanujan unitary Cayley graphs ICG$(n, \{1\})$. We extend Droll's result by drawing up a complete list of all graphs ICG$(p^s, \mathcal{D})$ having the Ramanujan property for each prime power p^s and arbitrary divisor set $\mathcal{D}$. In order to do this we shall have to check for which $\mathcal{D} \subseteq D^*(p^s)$ the Ramanujan condition

$$\Lambda(p^s, \mathcal{D}) := \Lambda(\mathrm{ICG}(p^s, \mathcal{D})) \leq 2\sqrt{\Phi(p^s, \mathcal{D}) - 1} \tag{4.30}$$

is satisfied (cf. Definition 2.34 and Proposition 4.1 (ii)). Since Ramanujan graphs are required to be connected, we necessarily have $1 \in \mathcal{D}$ for a Ramanujan graph ICG$(p^s, \mathcal{D})$ due to the criterion of So (cf. Proposition 4.1 (iii)). While (4.27) provides an explicit formula for $\Phi(p^s, \mathcal{D})$, it is just as important to have such a formula for $\Lambda(p^s, \mathcal{D})$. As a tool to prove such a formula we first show

Lemma 4.4. *Let p^s be a prime power and $\mathcal{D} = \{p^{a_1}, p^{a_2}, \ldots, p^{a_{r-1}}, p^{a_r}\}$ with integers $0 \leq a_1 < a_2 < \cdots < a_{r-1} < a_r \leq s$. Then*

$$\lambda_k(p^s, \mathcal{D}) = \sum_{\substack{i=1 \\ a_i \geq s-j}}^{r} \varphi(p^{s-a_i}) - \sum_{\substack{i=1 \\ a_i = s-j-1}}^{r} p^{s-a_i-1} \tag{4.31}$$

for $k \in \{1, 2, \ldots, p^s\}$, where $j := e_p(k)$.

Proof. By (4.26) in Proposition 4.1, we have for $k \in \{1, 2, \ldots, p^s\}$ that

$$\lambda_k(p^s, \mathcal{D}) = \sum_{d \in \mathcal{D}} c\Big(k, \frac{p^s}{d}\Big) = \sum_{i=1}^{r} c(k, p^{s-a_i}). \tag{4.32}$$

We use two well-known properties of Ramanujan sums (cf. [4] or [44]), namely $c(k,n) = c(\gcd(k,n), n)$ for all k and n, and

$$c(p^u, p^v) = \begin{cases} \varphi(p^v), & \text{if } u \geq v, \\ -p^{v-1}, & \text{if } u = v-1, \\ 0, & \text{if } u \leq v-2, \end{cases}$$

for primes p and non-negative integers u and v. On setting $m := \frac{k}{p^j}$, i.e. $k = p^j m$ with $0 \leq j \leq s$ and $m \geq 1$, $p \nmid m$, it follows that

$$\begin{aligned} c(k, p^{s-a_i}) &= c(p^j m, p^{s-a_i}) \\ &= c(p^{\min\{j, s-a_i\}}, p^{s-a_i}) = \begin{cases} \varphi(p^{s-a_i}), & \text{if } j \geq s-a_i, \\ -p^{s-a_i-1}, & \text{if } j = s-a_i-1, \\ 0, & \text{if } j \leq s-a_i-2. \end{cases} \end{aligned}$$

Inserting this into (4.32), we obtain (4.31). □

Exercise 4.5. Check the identities for Ramanujan sums stated in the preceding proof.

Proposition 4.6. *Let $p^s \geq 3$ be a prime power, and let $\mathcal{D} \subseteq D(p^s)$ with $1 \in \mathcal{D}$. Then $\Lambda(2^s, \{1\}) = 0$ for $s \geq 2$, and in all other cases*

$$\Lambda(p^s, \mathcal{D}) = p^{s-1} - \sum_{\substack{d \in \mathcal{D} \\ d \neq 1}} \varphi\Big(\frac{p^s}{d}\Big),$$

where the empty sum for $\mathcal{D} = \{1\}$ vanishes.

Proof. Since $1 \in \mathcal{D}$, we have $\mathcal{D} = \{p^{a_1}, p^{a_2}, \ldots, p^{a_r}\}$, say, for suitable integers $0 = a_1 < a_2 < \cdots < a_{r-1} < a_r \leq s$. Writing $k = p^j m$, $p \nmid m$, with suitable integers j, $0 \leq j \leq s$, and $m \geq 1$ for each $k \in \{1, 2, \ldots, p^s\}$, we obtain, by Lemma 4.4 (= 4.2 in [29]), that

$$\lambda_k(p^s, \mathcal{D}) = \sum_{\substack{i=1 \\ a_i \geq s-j}}^{r} \varphi(p^{s-a_i}) - \sum_{\substack{i=1 \\ a_i = s-j-1}}^{r} p^{s-a_i-1}. \tag{4.33}$$

We distinguish several cases. If $j = s$, that is $k = p^s$, we have $\lambda_k(p^s, \mathcal{D}) = \Phi(p^s, \mathcal{D})$, which is the largest eigenvalue of ICG$(p^s, \mathcal{D})$ (see (iii)) and thus irrelevant for the determination of $\Lambda(p^s, \mathcal{D})$. Hence we are left with the following three cases, the last two of which correspond with the cases in the proof of Proposition 4.1 in [29]:

Case 0: $0 \leq j \leq s - a_r - 2$.

Observe that this only occurs if $a_r \leq s - 2$. Then (4.33) implies

$$\lambda_k(p^s, \mathcal{D}) = 0.$$

Case 1: $s - a_\ell \leq j \leq s - a_{\ell-1} - 2$ for some $2 \leq \ell \leq r$.

Then

$$\lambda_k(p^s, \mathcal{D}) = \sum_{\substack{i=1 \\ a_i \geq a_\ell}}^{r} \varphi(p^{s-a_i}) = \Phi(p^s, \mathcal{D}(p^{a_\ell})) > 0$$

with $\mathcal{D}(x) := \{d \in \mathcal{D} : d \geq x\}$.

Case 2: $j = s - a_\ell - 1$ for some $1 \leq \ell \leq r$.

Then, on setting $a_{r+1} := s + 1$ and consequently $\Phi(p^s, \mathcal{D}(p^{a_{r+1}})) = 0$, we have

$$\begin{aligned}\lambda_k(p^s, \mathcal{D}) &= \sum_{\substack{i=1 \\ a_i \geq a_\ell+1}}^{r} \varphi(p^{s-a_i}) - p^{s-a_\ell-1} \\ &= \Phi(p^s, \mathcal{D}(p^{a_{\ell+1}})) - p^{s-a_\ell-1} \leq 0,\end{aligned}$$

such that $|\lambda_k(p^s, \mathcal{D})| = p^{s-a_\ell-1} - \Phi(p^s, \mathcal{D}(p^{a_{\ell+1}}))$.

Let us start by looking at the special case $r = 1$, i.e., $\mathcal{D} = \{1\}$, $a_1 = 0$ and $s \geq 1$. This means that Case 1 never occurs. Case 2 does occur for $j = s - 1$, with corresponding $\lambda_k(p^s, \mathcal{D}) = p^{s-1}$. This value equals $\Phi(2^s, \{1\}) = \varphi(2^s) = 2^{s-1}$ in case $p = 2$ (and $s \geq 2$, since $s = 1$ is outruled by our condition $p^s \geq 3$), hence $\Lambda(2^s, \{1\}) = 0$ (originating from Case 0). For any $p \geq 3$ and all $s \geq 1$, we have $p^{s-1} < \Phi(p^s, \{1\})$, which proves the proposition in this situation.

From now on, we assume that $r \geq 2$. We define

$$m = m_{\mathcal{D}} := \begin{cases} \infty, & \text{if } \mathcal{D} \text{ is uni-regular,} \\ \min\limits_{\substack{2 \leq \ell \leq r \\ a_\ell - a_{\ell-1} \geq 2}} \ell, & \text{otherwise.} \end{cases} \tag{4.34}$$

For $m_{\mathcal{D}} = \infty$, Case 1 does not occur at all. Obviously,

$$\Phi(p^s, \mathcal{D}) \geq \Phi(p^s, \mathcal{D}(p^{a_\ell})) > \Phi(p^s, \mathcal{D}(p^{a_{\ell+1}})) > 0$$

for $1 \leq \ell \leq r - 1$. If $m_{\mathcal{D}} < \infty$, Case 1 occurs for $\ell = m_{\mathcal{D}}$, but for no smaller ℓ. Therefore, the only candidate for $\Lambda(p^s, \mathcal{D})$ originating from Case 1 is $\Phi(p^s, \mathcal{D}(p^{a_m}))$. On the other hand, it is easily seen that

$$\Phi(p^s, \mathcal{D}) > p^{s-a_\ell-1} - \Phi(p^s, \mathcal{D}(p^{a_{\ell+1}})) \geq p^{s-a_{\ell+1}-1} - \Phi(p^s, \mathcal{D}(p^{a_{\ell+2}})) \geq 0$$

for $1 \leq \ell \leq r-1$. Hence, the only candidate for $\Lambda(p^s, \mathcal{D})$ originating from Case 2 is $p^{s-a_1-1} - \Phi(p^s, \mathcal{D}(p^{a_2})) = p^{s-1} - \Phi(p^s, \mathcal{D}(p^{a_2}))$. Now let us compare the two candidates for $\Lambda(p^s, \mathcal{D})$ found above. By the definition of $m = m_{\mathcal{D}}$, we obtain

$$\begin{aligned}
\Phi(p^s, \mathcal{D}(p^{a_2})) + \Phi(p^s, \mathcal{D}(p^{a_m})) &= \sum_{i=2}^{r} \varphi(p^{s-a_i}) + \sum_{i=m}^{r} \varphi(p^{s-a_i}) \\
&= \sum_{i=2}^{m-1} \varphi(p^{s-a_i}) + 2\sum_{i=m}^{r} \varphi(p^{s-a_i}) \\
&= p^{s-a_2} - p^{s-a_{m-1}-1} + 2\sum_{i=m}^{r} \varphi(p^{s-a_i}) \\
&\leq p^{s-a_2} - p^{s-a_{m-1}-1} + 2p^{s-a_m} \leq p^{s-1}.
\end{aligned}$$

This implies $p^{s-1} - \Phi(p^s, \mathcal{D}(p^{a_2})) \geq \Phi(p^s, \mathcal{D}(p^{a_m}))$, and we have

$$\Lambda(p^s, \mathcal{D}) = p^{s-1} - \Phi(p^s, \mathcal{D}(p^{a_2})) = p^{s-1} - \sum_{\substack{d \in \mathcal{D} \\ d \neq 1}} \varphi\Big(\frac{p^s}{d}\Big). \qquad \square$$

Corollary 4.7. *Let $p^s \geq 3$ be a prime power, and let $\mathcal{D} \subseteq D(p^s)$ with $1 \in \mathcal{D}$.*

(i) *Then*

$$\Phi(p^s, \mathcal{D}) + \Lambda(p^s, \mathcal{D}) = \begin{cases} 2^{s-1}, & \text{if } p = 2,\ s \geq 2 \text{ and } \mathcal{D} = \{1\}, \\ p^s, & \text{otherwise.} \end{cases} \tag{4.35}$$

(ii) *For all odd primes p, we have $\Lambda(p^s, \mathcal{D}) = 0$ if and only if $\mathcal{D} = D(p^s)$.*

Proof. The identity (4.35) follows right away from (4.27) and Proposition 4.6. The second assertion is another consequence of Proposition 4.6, because

$$\sum_{\substack{d \in \mathcal{D} \\ d \neq 1}} \varphi\Big(\frac{p^s}{d}\Big) \leq p^{s-1},$$

with equality if and only if $\mathcal{D} = D(p^s)$. □

Corollary 4.8. *Let $p^s \geq 3$ be a prime power, and let $\mathcal{D} \subseteq D^*(p^s)$ with $1 \in \mathcal{D}$. Then* ICG$(p^s, \mathcal{D})$ *is Ramanujan if and only if either $p = 2$, $s \geq 2$ and $\mathcal{D} = \{1\}$ or, in all other cases,*

$$\sum_{\substack{d \in \mathcal{D} \\ d \neq 1}} \varphi\Big(\frac{p^s}{d}\Big) \geq p^{s-1} - 2\sqrt{p^s} + 2, \tag{4.36}$$

where the empty sum for $\mathcal{D} = \{1\}$ vanishes.

Proof. The special case $p = 2$, $s \geq 2$ and $\mathcal{D} = \{1\}$ is an immediate consequence of Proposition 4.6. From (4.30) and Proposition 4.6 it follows in all other situations that $\mathrm{ICG}(p^s, \mathcal{D})$ is Ramanujan if and only if $\Phi_2 := \sum_{d \in \mathcal{D}, d \neq 1} \varphi(\frac{p^s}{d})$ satisfies

$$p^{s-1} - \Phi_2 \leq 2\sqrt{\Phi(p^s, \mathcal{D}) - 1} = 2\sqrt{\varphi(p^s) + \Phi_2 - 1}.$$

A short calculation reveals that this is equivalent with (4.36). □

Expanding our definition of $D(n)$, we set $D(n; m) := \{d \in D(n) : d \geq m\}$ for any positive integer m. Now we are able to show precisely which ICGs of prime power order do have the Ramanujan property.

Theorem 4.9 ([30]). *Let p be a prime and s a positive integer such that $p^s \geq 3$. Let $\mathcal{D} \subseteq D^*(p^s)$ be an arbitrary divisor set of p^s. Then $\mathrm{ICG}(p^s, \mathcal{D})$ is a Ramanujan graph if and only if $\mathcal{D}$ lies in one of the following classes:*

(i) *$\mathcal{D} = D(p^{\lceil \frac{s}{2} \rceil - 1}) \cup \mathcal{D}'$ for some $\mathcal{D}' \subseteq D(p^{s-1}; p^{\lceil \frac{s}{2} \rceil})$;*

(ii) *$\mathcal{D} = \{1\}$ in case $p = 2$ and $s \geq 3$;*

(iii) *$\mathcal{D} = D(p^{\frac{s-3}{2}}) \cup \mathcal{D}'$ such that $|\mathcal{D}| \geq 2$ for some $\mathcal{D}' \subseteq D(p^{s-1}; p^{\frac{s+1}{2}})$ in case $p \in \{2, 3\}$ and $s \geq 3$ odd;*

(iv) *$\mathcal{D} = D(2^{\frac{s-4}{2}}) \cup \mathcal{D}'$ for some $\mathcal{D}' \subseteq D(2^{s-1}; 2^{\frac{s}{2}})$ satisfying $\emptyset \neq \mathcal{D}' \neq \{2^{s-1}\}$ in case $p = 2$ and $s \geq 4$ even;*

(v) *$\mathcal{D} = \{1, 2^2, 2^3, 2^4\}$ in case $p = 2$ and $s = 5$;*

(vi) *$\mathcal{D} = D(5^{\frac{s-3}{2}}) \cup \{5^{\frac{s+1}{2}}\} \cup \mathcal{D}'$ for some $\mathcal{D}' \subseteq D(5^{s-1}; 5^{\frac{s+3}{2}})$ in case $p = 5$ and $s \geq 5$ odd;*

(vii) *$\mathcal{D} = D(2^{\frac{s-5}{2}}) \cup \{2^{\frac{s-1}{2}}\} \cup \mathcal{D}'$ for some $\mathcal{D}' \subseteq D(2^{s-1}; 2^{\frac{s+1}{2}})$ satisfying*

$$3 - 2\sqrt{2} + \frac{1}{2^{\frac{s-3}{2}}} \leq 2^{\frac{s-1}{2}} \sum_{d' \in \mathcal{D}'} \frac{1}{d'} \tag{4.37}$$

in case $p = 2$ and $s \geq 5$ odd.

Theorem 4.9 tells us for each prime power in a simple way how to choose divisor sets to obtain an integral circulant graph that is Ramanujan, except for the final case (vii), where the construction is a little more intricate. In this situation, the binary expansion of the left-hand side of (4.37) is obtained by adding $2^{-\frac{s-3}{2}}$ to the binary expansion of the real constant $3 - 2\sqrt{2} = (0.0010101111101100001\ldots)_2$. Obviously, the sum on the right-hand side of (4.37) is a binary expansion by construction, and in order to generate a Ramanujan graph we just have to pick $\mathcal{D}'$ appropriately. Let us sketch an example that illustrates what to do explicitly in case (vii):
Consider the prime power 2^{29}, for which condition (4.37) turns into

$$3 - 2\sqrt{2} + \frac{1}{2^{13}} = (0.0010101111110100001\ldots)_2 \leq \sum_{i=15}^{r} \frac{1}{2^{a_i - 14}},$$

with $\mathcal{D}' = \{2^{a_{15}}, 2^{a_{16}}, \dots, 2^{a_r}\}$, say. The parameter $r \geq 15$ may be selected arbitrarily so that $a_r \leq 28$. Now for each choice of the a_i, $15 \leq i \leq r$, it is completely obvious whether the corresponding divisor set $\mathcal{D}'$ generates a Ramanujan graph or not. For instance, the choice $a_{15} = 17$, $a_{16} = 19$, $a_{17} = 21$, $a_{18} = 22$, $a_{19} = 23$, $a_{20} = 24$, $a_{21} = 25$, $a_{22} = 26$, $a_{23} = 27$ generates the Ramanujan graph

$$\mathrm{ICG}(2^{29}, D(2^{12}) \cup \{2^{14}, 2^{17}, 2^{19}, 2^{21}, 2^{22}, 2^{23}, 2^{24}, 2^{25}, 2^{26}, 2^{27}\}),$$

while the divisor set of

$$\mathrm{ICG}(2^{29}, D(2^{12}) \cup \{2^{14}, 2^{17}, 2^{19}, 2^{21}, 2^{22}, 2^{23}, 2^{24}, 2^{25}, 2^{26}, 2^{28}\})$$

violates (4.37), and thus the graph is not Ramanujan.

Theorem 4.9 (i) immediately implies the following statement.

Corollary 4.10. *For each prime power* $p^s \geq 3$ *and all divisor sets* $\mathcal{D} = \{1, p, \dots, p^{r-1}\}$ *with* $\frac{s}{2} \leq r \leq s$, *the graph* $\mathrm{ICG}(p^s, \mathcal{D})$ *is Ramanujan. In particular, there is an integral circulant Ramanujan graph* $\mathrm{ICG}(p^s, \mathcal{D})$ *for each prime power* $p^s \geq 3$.

We like to make the reader aware of the fact that the divisor sets ensuring the Ramanujan property in Corollary 4.10 are *uni-regular*, as introduced in [29], where a subset of $D(p^s)$ is called uni-regular if it consists of successive powers of p.

Proof of Theorem 4.9. Let us assume that $\mathcal{D} = \{p^{a_1}, p^{a_2}, \dots, p^{a_r}\}$ for suitable integers $0 \leq a_1 < a_1 < a_2 < \cdots < a_{r-1} < a_r \leq s-1$. Since Ramanujan graphs have to be connected by definition, i.e., $1 \in \mathcal{D}$, we necessarily have $a_1 = 0$.

We consider the case $r = 1$, i.e., $\mathcal{D} = \{1\}$ separately. We know by Corollary 4.8 that $\mathrm{ICG}(2^s, \{1\})$ is Ramanujan for each $s \geq 2$. For $p \geq 3$, the Ramanujan property requires $p^{s-1} - 2p^{s/2} + 2 \leq 0$ by condition (4.36) of Corollary 4.8. It is easy to check that this inequality holds if and only if $s = 1$ or $s = 2$. For $r = 1$, we thus have exactly the following types of Ramanujan ICGs:

	p^s	$\mathcal{D}$
A	2^s $(s \geq 2)$	$\{1\}$
B	p $(p \geq 3)$	$\{1\}$
C	p^2 $(p \geq 3)$	$\{1\}$

In the sequel we may assume that $r \geq 2$. Since $a_r \leq s-1$, we trivially have

$$\sum_{i=2}^{r} \varphi(p^{s-a_i}) \leq (p^{s-a_2} - p^{s-a_2-1}) + (p^{s-a_2-1} - p^{s-a_2-2}) + \cdots + (p-1) = p^{s-a_2} - 1.$$

Hence, by Corollary 4.8, a necessary condition for $\mathrm{ICG}(p^s, \mathcal{D})$ to be Ramanujan is that

$$p^{s-a_2} \geq p^{s-1} - 2\sqrt{p^s} + 2. \tag{4.38}$$

Let us first distinguish cases according to the value of a_2. If $a_2 \geq 3$, thus $s \geq 4$, a simple calculation shows that (4.38) is never satisfied. Assume next that $a_2 = 2$, thus $s \geq 3$. It is easily seen that (4.38) requires $s \leq 5$, and we obtain only the pairs $(s, p) \in \{(3,2), (3,3), (4,2), (5,2)\}$ satisfying (4.38). Checking our Ramanujan condition (4.36) explicitly for these pairs, we precisely obtain Ramanujan ICGs as listed below:

	p^s	$\mathcal{D}$
D	$p^3,\ 2 \leq p \leq 3$	$\{1, p^2\}$
E	2^4	$\{1, 2^2\}$ or $\{1, 2^2, 2^3\}$
F	2^5	$\{1, 2^2, 2^3, 2^4\}$

We are left with the case $a_2 = 1$. Using the notation introduced in (4.34), it follows that $m_{\mathcal{D}} \geq 3$, and we shall distinguish between $m_{\mathcal{D}} = \infty$ and $m_{\mathcal{D}} < \infty$. In the first case, i.e. for uni-regular divisor sets $\mathcal{D}$, we have $s \geq 2$ and

$$\sum_{i=2}^{r} \varphi(p^{s-a_i}) = \sum_{i=2}^{r} \varphi(p^{s-(i-1)}) = \varphi(p^{s-1}) - \varphi(p^{s-r}).$$

Then Corollary 4.8 tells us that $\mathrm{ICG}(p^s, \mathcal{D})$ is Ramanujan if and only if

$$p^{s-r} + 2 \leq 2\sqrt{p^s}. \tag{4.39}$$

For $s \leq 2r$, this condition is satisfied for all primes p, hence we obtain a Ramanujan graph $\mathrm{ICG}(p^s, \mathcal{D})$ for each prime power p^s with $s \geq 2$ and uni-regular divisor sets $\mathcal{D} = \{1, p, \ldots, p^{r-1}\}$, where $\frac{s}{2} \leq r \leq s$. For $s \geq 2r + 2$, condition (4.39) does not hold for any prime p. For the missing case $s = 2r + 1$, (4.39) yields the Ramanujan condition $\sqrt{p} + \frac{2}{p^{s/2}} \leq 2$, which is only satisfied for $p = 2$ or $p = 3$ and all odd $s \geq 5$. Therefore, we have two more types of Ramanujan ICGs:

	p^s	$\mathcal{D}$	
G	$p^s\ (p \in \mathbb{P},\ s \geq 2)$	$\{1, p, \ldots, p^{r-1}\}$	$(\min\{2, \frac{s}{2}\} \leq r \leq s)$
H	$p^s\ (p \in \{2,3\},\ s \geq 5,\ 2 \nmid s)$	$\{1, p, \ldots, p^{\frac{s-3}{2}}\}$	

Now let us consider the case $m = m_{\mathcal{D}} < \infty$. We have

$$\begin{aligned}\sum_{i=2}^{r} \varphi(p^{s-a_i}) &= \sum_{i=2}^{m-1} \varphi(p^{s-a_i}) + \sum_{i=m}^{r} \varphi(p^{s-a_i}) \\ &= p^{s-1} - p^{s-a_{m-1}-1} + \sum_{i=m}^{r} \varphi(p^{s-a_i}).\end{aligned}$$

Then, by Corollary 4.8, $\mathrm{ICG}(p^s, \mathcal{D})$ is Ramanujan if and only if

$$p^{s-a_{m-1}-1} + 2 \leq 2\sqrt{p^s} + \sum_{i=m}^{r} \varphi(p^{s-a_i}). \tag{4.40}$$

For $s \leq 2(a_{m-1}+1)$, i.e., $\frac{s}{2} \geq s - a_{m-1} - 1$, the Ramanujan condition (4.40) is satisfied for all primes p. By the definition of $m = m_{\mathcal{D}}$, we have $a_{m-1} = m-2$. Hence $\mathrm{ICG}(p^s, \mathcal{D})$ is Ramanujan for any prime power p^s and $\mathcal{D} = \{1, p, \ldots, p^{m-2}, p^{a_m}, \ldots, p^{a_r}\}$, if $3 \leq m \leq r$ satisfies $s \leq 2(m-1)$ and $a_m \geq m$, which requires $s \geq 4$. Setting $\ell := m-1$, we have found the next type of Ramanujan ICGs:

	p^s	$\mathcal{D}$	
I	p^s ($p \in \mathbb{P}$, $s \geq 4$)	$\{1, p, \ldots, p^{\ell-1}, p^{a_{\ell+1}}, \ldots, p^{a_r}\}$	$(\frac{s}{2} \leq \ell \leq a_{\ell+1} - 1)$

For $s \geq 2(a_{m-1}+3)$, i.e. $\frac{s}{2} \leq s - a_{m-1} - 3$, we obtain for all primes p

$$p + \frac{2}{p^{s-a_{m-1}-1}} > 2 \geq \frac{p^{\frac{s}{2}+1}}{p^{s-a_{m-1}-2}} + 1 \geq \frac{2p^{\frac{s}{2}}}{p^{s-a_{m-1}-2}} + 1,$$

hence

$$\begin{aligned} p^{s-a_{m-1}-1} + 2 &= p^{s-a_{m-1}-2}\left(p + \frac{2}{p^{s-a_{m-1}-2}}\right) \\ &> p^{s-a_{m-1}-2}\left(\frac{2p^{\frac{s}{2}}}{p^{s-a_{m-1}-2}} + 1\right) = 2p^{\frac{s}{2}} + p^{s-a_{m-1}-2} \\ &\geq 2p^{\frac{s}{2}} + p^{s-a_m} > 2\sqrt{p^s} + \sum_{i=m}^{r} \varphi(p^{s-a_i}). \end{aligned}$$

This contradicts (4.40), which means that there are no Ramanujan graphs in this situation.

It remains to consider the three special cases $s = 2(a_{m-1}+1) + j$ with $j \in \{1, 2, 3\}$. This turns (4.40) into the Ramanujan condition

$$p^{\frac{j}{2}} - 2 \leq \frac{1}{p^{a_{m-1}+1+\frac{j}{2}}}\left(\sum_{i=m}^{r} \varphi(p^{s-a_i}) - 2\right). \tag{4.41}$$

By the definition of $m = m_{\mathcal{D}}$, we know that $a_m \geq a_{m-1} + 2$. Since

$$\sum_{i=m}^{r} \varphi(p^{s-a_i}) \leq p^{s-a_m} - 1 \leq p^{2(a_{m-1}+1)+j-(a_{m-1}+2)} - 1 = p^{a_{m-1}+j} - 1,$$

careful calculations reveal that condition (4.41) can only be satisfied for the pairs $(j, p) \in \{(1,2), (1,3), (1,5), (2,2), (3,2)\}$. For the first two pairs, corresponding with $a_{m-1} = \frac{s-3}{2}$, the left-hand side of (4.41) is negative, while the right-hand side is non-negative except for the special case $p = 2$ and $a_m = a_r = s - 1$. However, a closer look shows that in this latter situation (4.41) is still satisfied, and we obtain Ramanujan graphs $\mathrm{ICG}(p^s, \mathcal{D})$ for $p \in \{2,3\}$, where $\mathcal{D} = \{1, p, \ldots, p^{\frac{s-3}{2}}, p^{a_{\frac{s+1}{2}}}, \ldots, p^{a_r}\}$ with $a_{\frac{s+1}{2}} = a_m \geq a_{m-1} + 2 = \frac{s-3}{2} + 2 = \frac{s+1}{2}$, thus $s \geq 5$. Inserting the third pair $j = 1$ and $p = 5$ into (4.41), a short computation discloses that the corresponding ICGs are Ramanujan if and only if $a_m = a_{m-1} + 2$, and then $\mathcal{D} = \{\{1, 5, \ldots, 5^{\frac{s-3}{2}}, 5^{\frac{s+1}{2}}, \ldots, 5^{a_r}\}$ with $s \geq 5$. To sum up, the Ramanujan ICGs for $(j, p) \in \{(1,2), (1,3), (1,5)\}$ are the following:

	p^s	$\mathcal{D}$	
J	p^s $(p \in \{2,3\},\ s \geq 5,\ 2 \nmid s)$	$\{1, p, \ldots, p^{\frac{s-3}{2}}, p^{a_{\frac{s+1}{2}}}, \ldots, p^{a_r}\}$	$(a_{\frac{s+1}{2}} \geq \frac{s+1}{2})$
K	5^s $(s \geq 5,\ 2 \nmid s)$	$\{1, 5, \ldots, 5^{\frac{s-3}{2}}, 5^{\frac{s+1}{2}}, 5^{a_{\frac{s+3}{2}}}, \ldots, 5^{a_r}\}$	

For $j = p = 2$, the Ramanujan condition (4.41) reads $\sum_{i=m}^{r} \varphi(p^{s-a_i}) \geq 2$, and this always holds unless $a_m = a_r = s - 1$, i.e., we obtain Ramanujan ICGs of type

	p^s	$\mathcal{D}$	
L	2^s $(s \geq 6,\ 2 \mid s)$	$\{1, 2, \ldots, 2^{\frac{s-4}{2}}, 2^{a_{\frac{s}{2}}}, \ldots, 2^{a_r}\}$	$(\frac{s}{2} \leq a_{\frac{s}{2}} \leq s - 2)$

We are left with the case $j = 3$ and $p = 2$, where $s = 2a_{m-1} + 5 = 2(m-2) + 5 = 2m + 1$, and our Ramanujan condition (4.41) becomes

$$
\begin{aligned}
2\sqrt{2} - 2 &\leq \frac{1}{2^{a_{m-1}+\frac{5}{2}}} \left(\sum_{i=m}^{r} \varphi(2^{s-a_i}) - 2 \right) \\
&= \frac{1}{2^m \sqrt{2}} \left(\sum_{i=m}^{r} 2^{s-a_i-1} - 2 \right) = \frac{1}{\sqrt{2}} \left(\sum_{i=m}^{r} \frac{1}{2^{a_i - m}} - \frac{1}{2^{m-1}} \right).
\end{aligned}
$$

Apparently, this is equivalent to $4 - 2\sqrt{2} \leq \sum_{i=m}^{r} \frac{1}{2^{a_i-m}} - \frac{1}{2^{m-1}}$, where we know that $a_i \geq i$ for all $i \geq m$. This immediately implies that the Ramanujan condition necessarily requires that $a_m = m$. Then our condition reads

$$
3 - 2\sqrt{2} \leq \sum_{i=m+1}^{r} \frac{1}{2^{a_i - m}} - \frac{1}{2^{m-1}} = \sum_{i=\frac{s+1}{2}}^{r} \frac{1}{2^{a_i - \frac{s-1}{2}}} - \frac{1}{2^{\frac{s-3}{2}}}, \tag{4.42}
$$

and we have found our last type of Ramanujan ICGs:

	p^s	$\mathcal{D}$
M	2^s $(s \geq 5,\ 2 \nmid s)$	$\{1, 2, \ldots, 2^{\frac{s-5}{2}}, 2^{\frac{s-1}{2}}, 2^{a_{\frac{s+1}{2}}}, \ldots, 2^{a_r}\}$ $(a_{\frac{s+1}{2}}, \ldots, a_r$ satisfy (4.42))

To complete the proof of Theorem 4.9, it remains to show that the Ramanujan ICGs of types A–M exactly constitute the Ramanujan graphs listed in the assertion. The following table provides the interrelations between types A–M on one hand and cases (i)–(vii) on the other hand.

(i)	A $(s = 2)$, B $(s = 1)$, C $(s = 2)$, G (s$\geq$ 2), I $(s \geq 4)$
(ii)	A $(s \geq 3)$
(iii)	D $(s = 3)$, H $(s \geq 5$ odd), J $(s \geq 5$ odd)
(iv)	E $(s = 4)$, L $(s \geq 6$ even)
(v)	F $(s = 5)$
(vi)	K $(s \geq 5$ odd)
(vii)	M $(s \geq 5$ odd)

□

Applying the concept of multiplicative divisor sets introduced at the end of Section 4.1, Corollary 4.10 can be extended to integral circulant graphs whose order is an arbitrary composite number. It has recently been shown that for every even integer $n \geq 4$ and a positive proportion of the odd integers n, namely those having a "dominating" prime power factor, there exists a divisor set $\mathcal{D} \subseteq D^*(n)$ such that $\mathrm{ICG}(n, \mathcal{D})$ is Ramanujan. Unfortunately, the corresponding assertion cannot be confirmed for all odd n by use of the tool of multiplicative divisor sets. In fact, it will transpire that the set of odd n for which no Ramanujan graph $\mathrm{ICG}(n, \mathcal{D})$ with multiplicative divisor set $\mathcal{D}$ exists has positive density. This means that one would definitely need to study integral circulant graphs with non- multiplicative divisor sets in order to finally settle the matter.

At the end of this section we just state and discuss some further results and finally ask a couple of questions.

Theorem 4.11 ([39] (2013)). *For each even integer $n \geq 3$ there is a* (*multiplicative*) *divisor set $\mathcal{D} \subseteq D^*(n)$ such that* $\mathrm{ICG}(n, \mathcal{D})$ *is a Ramanujan graph.*

If n is an odd integer, a multiplicative divisor set $\mathcal{D} \subseteq D^*(n)$ such that $\mathrm{ICG}(n, \mathcal{D})$ is Ramanujan can only exist if n has a comparatively large prime power factor. This will be specified in detail by the following theorem. To this end, we define for every integer $n > 1$ its largest prime power factor $\mathrm{PP}(n) := \max_{p \in \mathbb{P},\, p \mid n} p^{e_p(n)}$, and we shall say that n has a *dominating* prime power factor if $\mathrm{PP}(n) \geq \frac{n^{3/2}+n}{2(n-1)}$.

Theorem 4.12 ([39] (2013)). *Let $n \geq 3$ be an odd integer.*

(i) *If n has a dominating prime power factor, then there is a (multiplicative) divisor set $\mathcal{D} \subseteq D^*(n)$ such that* $\mathrm{ICG}(n, \mathcal{D})$ *is a Ramanujan graph.*

(ii) *If $n \geq 8259$ does not have a dominating prime power factor, then there is no multiplicative divisor set $\mathcal{D} \subseteq D^*(n)$ such that* $\mathrm{ICG}(n, \mathcal{D})$ *is a Ramanujan graph.*

There are a little less than 200 odd integers $n < 8259$ without dominating prime power factor, the smallest being $315 = 3^2 \cdot 5 \cdot 7$ and the largest $8211 = 3 \cdot 7 \cdot 17 \cdot 23$. For each of these n we assured ourselves that $\mathrm{ICG}(n, \mathcal{D})$ is not Ramanujan for the "canonical" multiplicative candidate $\mathcal{D} := \prod_{p \in \mathbb{P},\, p \mid n} \mathcal{D}_p$ to produce a Ramanujan integral circulant graph, namely by choosing $\mathcal{D}_q = D^*(q^{e_q(n)})$ for the largest prime power $q^{e_q(n)} = \mathbb{PP}(n)$ of n and $\mathcal{D}_p = D(p^{e_p(n)})$ for all other primes $p \mid n$. Checking, however, all possible multiplicative divisor sets for all those n would require an enormous computational effort, which is not justified by whatever the result is.

While Theorem 4.11 confirms the existence of a Ramanujan graph $\mathrm{ICG}(n, \mathcal{D})$ for all even n, Theorem 4.12 leaves unanswered the question if for every odd n there is a Ramanujan integral circulant graph of order n. However, Theorem 4.12 (ii) does tell us that in order to deal with this problem one would have to study integral circulant graphs with *non-multiplicative* divisor sets $\mathcal{D}$ for infinitely many n. Our final result discloses the proportion between the odd n for which $\mathrm{ICG}(n, \mathcal{D})$ is Ramanujan for some multiplicative divisor set $\mathcal{D} \subseteq D^*(n)$ and those odd n without this property. More precisely, it turns out that both cases occur with positive density among all odd integers.

For this purpose, let $\kappa(x)$ denote

$$\#\{n \leq x :\ 2 \nmid n,\ \mathrm{ICG}(n, \mathcal{D}) \text{ Ramanujan for some multiplicative } \mathcal{D} \subseteq D^*(n)\}.$$

It can be shown that $\lim_{x \to \infty} \frac{\kappa(x)}{x}$ exists and is positive. Notice that the set of odd prime powers below x, for each of which there is some Ramanujan ICG by Corollary 1.1 in [30] (see also Proposition 4.10), does not suffice to warrant this result, since it has density zero in the set of positive integers. By use of methods and results from analytic number theory, one can show

Theorem 4.13 ([39] (2013)). *For $x \to \infty$,*

$$\kappa(x) = \frac{\log 2}{2} x + O\Big(\frac{x}{\log x}\Big).$$

Corollary 4.14. *We have*

$$\lim_{x \to \infty} \frac{\kappa(x)}{\frac{x}{2}} = \log 2,$$

i.e., the density of odd positive integers n for which $\mathrm{ICG}(n, \mathcal{D})$ *is Ramanujan for some multiplicative divisor set $\mathcal{D} \subseteq D^*(n)$ in the set of all odd positive integers equals* $\log 2$, *while the corresponding density of odd positive integers n such that no* $\mathrm{ICG}(n, \mathcal{D})$ *with multiplicative $\mathcal{D}$ is Ramanujan equals* $1 - \log 2$.

Open problems:

- Given a (sufficiently large) positive integer n, is there some $\mathcal{D} \subseteq D^*(n)$ such that $\mathrm{ICG}(n, \mathcal{D})$ is Ramanujan?
- How do non-multiplicative divisor sets such that the corresponding integral circulant graphs are Ramanujan look?

4.3 The energy of a graph

In 1978 Gutman [15] introduced the mathematical concept of the *energy*

$$E(\mathcal{G}) := \sum_{\lambda \in \mathrm{Spec}(\mathcal{G})} |\lambda|.$$

of a graph $\mathcal{G}$, though this concept is rooted in chemistry way back in the 1930s (see [32] for connections between *Hückel molecular orbital theory* and graph spectral analysis, and [9] for a mathematical survey).

We denote the energy of an integral circulant graph by

$$\mathcal{E}(n, \mathcal{D}) := E(\mathrm{ICG}(n, \mathcal{D})) = \sum_{\lambda \in \mathrm{Spec}(\mathrm{ICG}(n,\mathcal{D}))} |\lambda| = \sum_{k=1}^{n} |\lambda_k(n, \mathcal{D})|. \tag{4.43}$$

It is of particular interest to determine for any fixed positive integer n the extremal energies

$$\mathcal{E}_{\min}(n) := \min\{\mathcal{E}(n, \mathcal{D}) :\ \mathcal{D} \subseteq D^*(n)\}$$

and

$$\mathcal{E}_{\max}(n) := \max\{\mathcal{E}(n, \mathcal{D}) :\ \mathcal{D} \subseteq D^*(n)\},$$

as well as the divisor sets producing these energies. For prime powers $n = p^s$ this problem was completely settled by the author and T. Sander in [40], Theorem 3.1, and [41], Theorem 1.1, respectively. The basis of these results was an explicit formula evaluating $\mathcal{E}(p^s, \mathcal{D})$ (cf. [40], Theorem 2.1). Due to the lack of a comparable energy formula for arbitrary n, it seems much more difficult to deal with $\mathcal{E}_{\min}(n)$ and $\mathcal{E}_{\max}(n)$ in general. Unfortunately, the energy reveals no sign of multiplicativity with respect to n. We shall overcome that deficiency by using the concept of *multiplicative divisor sets* (cf. Section 4.1 and [28]), which is closely linked with the eigenvalue representation in (4.28). This will at least provide us with good bounds for $\mathcal{E}_{\min}(n)$ and $\mathcal{E}_{\max}(n)$ as well as divisor sets producing the corresponding energies. In case of the minimal energy, we even conjecture to have found $\mathcal{E}_{\min}(n)$ and its associated divisor sets for all n.

By use of (4.28) and Proposition 4.3, Le and the author proved in [28], Theorem 4.2, that given a multiplicative divisor set $\mathcal{D} \subseteq D(n)$, we have for the energy of $\mathrm{ICG}(n, \mathcal{D})$, as defined in (4.43), that

$$\mathcal{E}(n, \mathcal{D}) = \prod_{p \in \mathbb{P},\, p|n} \mathcal{E}(p^{e_p(n)}, \mathcal{D}_p), \tag{4.44}$$

where the sets $\mathcal{D}_p$ are defined as in Proposition 4.3. We shall investigate minimal and maximal energies of integral circulant graphs with respect to multiplicative divisor sets. To this end, we define

$$\tilde{\mathcal{E}}_{\min}(n) := \min\{\mathcal{E}(n,\mathcal{D}) :\ \mathcal{D} \subseteq D^*(n) \text{ multiplicative}\}$$

and

$$\tilde{\mathcal{E}}_{\max}(n) := \max\{\mathcal{E}(n,\mathcal{D}) :\ \mathcal{D} \subseteq D^*(n) \text{ multiplicative}\}.$$

Clearly,

$$\mathcal{E}_{\min}(n) \le \tilde{\mathcal{E}}_{\min}(n) \le \tilde{\mathcal{E}}_{\max}(n) \le \mathcal{E}_{\max}(n) \tag{4.45}$$

for all positive integers n. Moreover, for prime powers p^s any divisor set $\mathcal{D} \subseteq D(p^s)$ is trivially multiplicative, hence $\tilde{\mathcal{E}}_{\min}(p^s) = \mathcal{E}_{\min}(p^s)$ and $\tilde{\mathcal{E}}_{\max}(p^s) = \mathcal{E}_{\max}(p^s)$.

We shall first state all our results and provide the corresponding proofs later.

Theorem 4.15 ([29] (2012)). *Let $n \ge 2$ be an integer with prime factorisation $n = p_1^{s_1} \cdot \dots \cdot p_t^{s_t}$. Then*

$$\tilde{\mathcal{E}}_{\max}(n) = \prod_{i=1}^{t} \theta(p_i^{s_i}),$$

where for any prime power p^s

$$\theta(p^s) := \begin{cases} \frac{1}{(p+1)^2}\big((s+1)(p^2-1)p^s + 2(p^{s+1}-1)\big), & \text{if } 2 \nmid s,\\ \frac{1}{(p+1)^2}\big(s(p^2-1)p^s + 2(2p^{s+1} - p^{s-1} + p^2 - p - 1)\big), & \text{if } 2 \mid s. \end{cases}$$

Moreover, for multiplicative sets $\mathcal{D} \subseteq D^*(n)$, we have $\mathcal{E}(n,\mathcal{D}) = \tilde{\mathcal{E}}_{\max}(n)$ if and only if $\mathcal{D} = \prod_{i=1}^{t} \mathcal{D}^{(i)}$ with

$$\mathcal{D}^{(i)} = $$

$$\begin{cases} \{1, p_i^2, p_i^4, \dots, p_i^{s_i-3}, p_i^{s_i-1}\}, & \text{if } 2 \nmid s_i,\ p_i \ge 3,\\ \{1, 2^2, 2^4, \dots, 2^{s_i-3}, 2^{s_i-1}\} \text{ or } \{1, 2, 2^3, \dots, 2^{s_i-4}, 2^{s_i-2}, 2^{s_i-1}\}, & \text{if } 2 \nmid s_i,\ p_i = 2,\\ \{1, p_i^2, p_i^4, \dots, p_i^{s_i-4}, p_i^{s_i-2}, p_i^{s_i-1}\} \text{ or } \{1, p_i, p_i^3, \dots, p_i^{s_i-3}, p_i^{s_i-1}\}, & \text{if } 2 \mid s_i. \end{cases}$$

It should be observed that the maximising factor divisor sets $\mathcal{D}^{(i)}$ are (almost) semi-regular (cf. (ii) at the end of Section 4.1).

In order to describe the multiplicative divisor sets minimising the energy, let us call $\mathcal{D} \subseteq D(p^s)$ for a prime power p^s *uni-regular*, if $\mathcal{D} = \{p^u, p^{u+1}, \dots, p^{v-1}, p^v\}$ for some integers $0 \le u \le v \le s$, which is a special case of semi-regularity if $u = 0$ and $v = s$.

Theorem 4.16 ([29] (2012)). *Let $n \geq 2$ be an integer with prime factorisation $n = p_1^{s_1} \cdot \dots \cdot p_t^{s_t}$ and $p_1 < p_2 < \dots < p_t$. Then*

$$\tilde{\mathcal{E}}_{\min}(n) = 2n\Big(1 - \frac{1}{p_1}\Big).$$

Moreover, for multiplicative sets $\mathcal{D} \subseteq D^*(n)$, we have $\mathcal{E}(n, \mathcal{D}) = \tilde{\mathcal{E}}_{\min}(n)$ if and only if $\mathcal{D} = \prod_{i=1}^{t} \mathcal{D}^{(i)}$ with $\mathcal{D}^{(1)} = \{p_1^u\}$ for some $u \in \{0, 1, \dots, s_1 - 1\}$ and arbitrary uni-regular sets $\mathcal{D}^{(i)}$ with $p_i^{s_i} \in \mathcal{D}^{(i)}$ for $2 \leq i \leq t$.

Remarks 4.17.

(i) According to Proposition 4.1 (iii), ICG$(n, \mathcal{D})$ is connected if and only if the elements of $\mathcal{D}$ are co prime. Assuming connectivity in Theorem 4.16, minimising sets $\mathcal{D} = \prod_{i=1}^{t} \mathcal{D}^{(i)}$ then necessarily have $\mathcal{D}^{(1)} = \{1\}$ and $\mathcal{D}^{(i)} = \{1, p_i, p_i^2, \dots, p_i^{s_i}\}$ for $2 \leq i \leq t$, i.e., all $\mathcal{D}^{(i)}$ with $2 \leq i \leq t$ have to be semi-regular sets of type 1.

(ii) The proof of Theorem 4.16 reveals that for $\mathcal{D} \subseteq D(n)$, i.e., possibly $n \in \mathcal{D}$, we would have $\tilde{\mathcal{E}}_{\min}(n) = n$ with minimising sets $\mathcal{D} = \prod_{i=1}^{t} \mathcal{D}^{(i)}$, where each $\mathcal{D}^{(i)}$ is an arbitrary uni-regular set containing $p_i^{s_i}$.

(iii) One could prove Theorem 4.16 by using the energy formula for ICG$(p^s, \mathcal{D})$ containing loops (cf. [28], Proposition 5.1). Instead we shall take a closer look at the second largest $|\lambda|$ with $\lambda \in$ Spec(ICG$(n, \mathcal{D})$). This provides more insight into the underlying structure.

As a consequence of (4.45), we immediately obtain from Theorems 4.15 and 4.16 the desired bounds for the extremal energies of integral circulant graphs with arbitrary divisor sets.

Corollary 4.18. *Let $n \geq 2$ be an integer with prime factorisation $n = p_1^{s_1} \cdot \dots \cdot p_t^{s_t}$ and $p_1 < p_2 < \dots < p_t$. Then*

(i) $\mathcal{E}_{\max}(n) \geq \tilde{\mathcal{E}}_{\max}(n) = \prod_{i=1}^{t} \theta(p_i^{s_i})$;

(ii) $\mathcal{E}_{\min}(n) \leq \tilde{\mathcal{E}}_{\min}(n) = 2n\big(1 - \frac{1}{p_1}\big)$.

Examples show that, while equality between $\mathcal{E}_{\max}(n)$ and $\tilde{\mathcal{E}}_{\max}(n)$ does occur occasionally, we usually have $\mathcal{E}_{\max}(n) > \tilde{\mathcal{E}}_{\max}(n)$ (cf. Example 4.20). This phenomenon is due to the fact that maximising divisor sets normally are not multiplicative. Yet $\tilde{\mathcal{E}}_{\max}(n)$ falls short of $\mathcal{E}_{\max}(n)$ by less than a comparatively small factor. To state this result, we denote by φ Euler's totient function, by $\tau(n)$ the number of positive divisors of n, and by $\omega(n)$ the number of distinct prime factors of n. As before, $e_p(n)$ denotes the order of the prime p in n.

Theorem 4.19 ([29] (2012)). *Let n be a positive integer. Then*

(i) $$\mathcal{E}_{\max}(n) \le n \sum_{d|n} \frac{\varphi(d)\tau(d)}{d}$$
$$= n \prod_{p\in\mathbb{P},\, p|n} \left(\tfrac{1}{2}\left(1-\tfrac{1}{p}\right)(e_p(n)+1)(e_p(n)+2)+\tfrac{1}{p}\right);$$

(ii) $\mathcal{E}_{\max}(n) < \left(\tfrac{3}{4}\right)^{\omega(n)} n\, \tau(n)^2$;

(iii) $\mathcal{E}_{\max}(n) \le \tilde{\mathcal{E}}_{\max}(n)\, \tau(n)$.

The proof of Theorem 4.19 (ii) will show that

$$\left(\tfrac{1}{4}\right)^{\omega(n)} \tau(n)^2 < \sum_{d|n} \frac{\varphi(d)\tau(d)}{d} < \left(\tfrac{3}{4}\right)^{\omega(n)} \tau(n)^2, \tag{4.46}$$

and (ii) is deduced from (i) by use of the upper bound in (4.46). We like to point out that the constants $\frac{1}{4}$ and $\frac{3}{4}$ in the lower and upper bound of (4.46), respectively, cannot be improved for all n, although we expect $\sum_{d|n} \frac{\varphi(d)\tau(d)}{d} \approx \left(\frac{1}{2}\right)^{\omega(n)} \tau(n)^2$ for most n. Yet, each of the bounds is sharp for integers with certain arithmetic properties (cf. Remark 1).

As far as the magnitude of $\mathcal{E}_{\max}(n)$ is concerned, it is well known that $\tau(n) = O(n^{\varepsilon})$ for any real $\varepsilon > 0$. In fact the true maximal order of $\tau(n)$ is approximately $n^{\frac{\log 2}{\log\log n}}$. On average, $\tau(n)$ is of order $\log n$, but for almost all n it is considerably smaller, because the normal order of $\tau(n)$ is roughly $(\log n)^{\log 2} \approx (\log n)^{0.693}$. For all these results as well as for the average and normal order of $\omega(n)$ the reader is referred to [17], §§ 18.1–18.2, § 22.13, and § 22.11.

As opposed to sets maximising the energy of integral circulant graphs, numerical calculations suggest that divisor sets minimising the energy are always multiplicative. According to that, equality in Corollary 4.18 (ii) should hold for all n. We shall specify this observation in two conjectures at the end of this section.

Now we start to prove all results stated so far in this section.

Proof of Theorem 4.15. Let $n = p_1^{s_1} \cdot \dots \cdot p_t^{s_t}$ be fixed, and let $\mathcal{D} \subseteq D^*(n)$ be a multiplicative set satisfying $\mathcal{E}(n,\mathcal{D}) = \tilde{\mathcal{E}}_{\max}(n)$. By (4.44) (cf. [28], Theorem 4.2),

$$\tilde{\mathcal{E}}_{\max}(n) = \mathcal{E}(n,\mathcal{D}) = \prod_{p\in\mathbb{P},\, p|n} \mathcal{E}(p^{e_p(n)},\mathcal{D}_p) = \prod_{i=1}^{t} \mathcal{E}(p_i^{s_i},\mathcal{D}_{p_i}).$$

This implies that $\mathcal{E}(p^{s_i},\mathcal{D}_{p_i}) = \mathcal{E}_{\max}(p_i^{s_i})$ for $1 \le i \le t$. From [41], Theorem 1.1, we conclude that $\mathcal{E}_{\max}(p_i^{s_i}) = \theta(p_i^{s_i})$ for all i, and the only corresponding divisor sets $\mathcal{D}^{(i)}$ are just the ones listed in our assertion. □

While equality between $\mathcal{E}_{\max}(n)$ and $\tilde{\mathcal{E}}_{\max}(n)$ does occur occasionally, we usually have $\mathcal{E}_{\max}(n) > \tilde{\mathcal{E}}_{\max}(n)$. This is illustrated by

Example 4.20. We have $\theta(p) = 2(p-1)$ for each prime p. By use of Theorem 4.15, we easily obtain $\tilde{\mathcal{E}}_{\max}(6) = \theta(2)\cdot\theta(3) = 8$, $\tilde{\mathcal{E}}_{\max}(105) = \theta(3)\cdot\theta(5)\cdot\theta(7) = 384$, and $\tilde{\mathcal{E}}_{\max}(21) = \theta(3)\cdot\theta(7) = 48$. Numerical evaluation of (4.43) yields $\mathcal{E}_{\max}(6) = 10$, $\mathcal{E}_{\max}(105) = 520$, but $\mathcal{E}_{\max}(21) = 48 = \tilde{\mathcal{E}}_{\max}(21)$.

In order to be able to compare $\mathcal{E}_{\max}(n)$ and $\tilde{\mathcal{E}}_{\max}(n)$ and finally prove Theorem 4.19 completely, we start by establishing an upper bound for $\mathcal{E}_{\max}(n)$.

Proof of Theorem 4.19 (i) & (ii). (i) By (4.43), (4.28), and the well-known Hölder identity for Ramanujan sums (cf. [4], chapt. 8.3-8.4), we have for any n and any divisor set $\mathcal{D} \subseteq D(n)$

$$\mathcal{E}(n,\mathcal{D}) = \sum_{k=1}^{n} \left| \sum_{d\in\mathcal{D}} \mu\Big(\frac{n}{(n,kd)}\Big) \frac{\varphi(\frac{n}{d})}{\varphi\big(\frac{n}{(n,kd)}\big)} \right|,$$

where μ denotes the Möbius function. Let $n/\mathcal{D} := \{\frac{n}{d} : d \in \mathcal{D}\} \subseteq D(n)$ be the set of complementary divisors of all $d \in \mathcal{D}$ with respect to n. Then

$$\begin{aligned}\mathcal{E}(n,\mathcal{D}) &= \sum_{k=1}^{n} \left| \sum_{d\in n/\mathcal{D}} \mu\Big(\frac{d}{(d,k)}\Big) \frac{\varphi(d)}{\varphi\big(\frac{d}{(d,k)}\big)} \right| \\ &\leq \sum_{k=1}^{n} \sum_{d\in n/\mathcal{D}} \frac{\varphi(d)}{\varphi\big(\frac{d}{(d,k)}\big)} = \sum_{d\in n/\mathcal{D}} \varphi(d) \sum_{g|d} \frac{1}{\varphi(\frac{d}{g})} \sum_{\substack{k=1\\(k,d)=g}}^{n} 1\,.\end{aligned}$$

Since

$$\sum_{\substack{k=1\\(k,d)=g}}^{n} 1 = \frac{n}{d} \sum_{\substack{r=1\\(r,d)=g}}^{d} 1 = \frac{n}{d} \sum_{\substack{r=1\\(r,\frac{d}{g})=1}}^{\frac{d}{g}} 1 = \frac{n}{d}\cdot\varphi\Big(\frac{d}{g}\Big),$$

we conclude that

$$\mathcal{E}(n,\mathcal{D}) \leq n \sum_{d\in n/\mathcal{D}} \frac{\varphi(d)\tau(d)}{d} \leq n \sum_{d|n} \frac{\varphi(d)\tau(d)}{d}\,.$$

Since this holds for all n, the inequality in (i) follows. Since φ, τ and the identity function id are all multiplicative, this property is carried over to $\frac{\varphi\cdot\tau}{\mathrm{id}}$ and its summatory function

$$f(n) := \sum_{d|n} \frac{\varphi(d)\tau(d)}{d} \qquad (n \in \mathbb{N}). \tag{4.47}$$

Given a prime power p^s, we obviously have

$$\begin{aligned} f(p^s) &= \sum_{j=0}^{s} \frac{\varphi(p^j)\cdot\tau(p^j)}{p^j} = 1 + \sum_{j=1}^{s} \frac{(p^j - p^{j-1})(j+1)}{p^j} \\ &= \frac{1}{2}\Big(1 - \frac{1}{p}\Big)(s+1)(s+2) + \frac{1}{p}. \end{aligned} \tag{4.48}$$

This completes the proof of (i).

(ii) Taking into account (i), it suffices to prove the upper bound in (4.46), but in order to justify the remark following Theorem 4.19 we shall verify the lower bound as well. By (4.48), we have for all primes p and all positive integers s that

$$\frac{1}{4} \le \frac{1}{2}\left(1 - \frac{1}{p}\right) < \frac{f(p^s)}{(s+1)^2} \le \frac{3}{4}\left(1 - \frac{2}{3p}\right) < \frac{3}{4}, \tag{4.49}$$

thus $\frac{1}{4}\tau(p^s)^2 < f(p^s) < \frac{3}{4}\tau(p^s)^2$. By the multiplicativity of $\tau(n)$ and the additivity of $\omega(n)$, which implies the multiplicativity of $c^{\omega(n)}$ for any positive constant c, this proves (4.46). □

Remark 1. It is easy to see that $\frac{f(p^s)}{(s+1)^2}$ comes close to the lower bound $\frac{1}{4}$ in (4.49) for $p = 2$ and large s and close to the upper bound $\frac{3}{4}$ for large p and $s = 1$. Hence, the lower bound in (4.46) is approached for integers n that are high powers of 2, while the upper bound is approached for squarefree integers n having large prime factors.

Proposition 4.21. *Let p be a prime, and let s be a positive integer. Then*

$$g(p^s) := \frac{p^s f(p^s)}{\tilde{\mathcal{E}}_{\max}(p^s)} \le \begin{cases} \frac{p+1}{2p}(s+2) & \text{for } 2 \nmid s, \\ \frac{p+1}{2p} \cdot \frac{(s+1)(s+2)}{s} & \text{for } 2 \mid s \end{cases}$$

for the function f defined in (4.47). More precisely, we have in particular $g(2) = 2$, $g(p^2) \le 3$ and $g(2^4) = \frac{64}{17}$.

Proof. By [41], Theorem 1.1, we know that $\tilde{\mathcal{E}}_{\max}(p^s) = \mathcal{E}_{\max}(p^s) = \theta(p^s)$ as defined in Theorem 4.15.

Case 1: $2 \nmid s$.

By use of (4.48) we obtain

$$\begin{aligned} g(p^s) &\le \frac{\Big(\frac{1}{2}\big(1 - \frac{1}{p}\big)(s+1)(s+2) + \frac{1}{p}\Big)(p+1)^2}{(s+1)(p^2-1) + 2p - \frac{2}{p^s}} \\ &\le \frac{\big(1 - \frac{1}{p}\big)(s+1)(s+2)(p+1)^2}{2(s+1)(p^2-1)} = \frac{p+1}{2p}(s+2). \end{aligned}$$

Observe that the second inequality is an identity for $p = 2$ and $s = 1$.

Case 2: $2 \mid s$.

Applying again (4.48), we conclude

$$\begin{aligned} g(p^s) &\le \frac{\left(\frac{1}{2}(1-\frac{1}{p})(s+1)(s+2)+\frac{1}{p}\right)(p+1)^2}{s(p^2-1)+4p-\frac{2}{p}+\frac{2}{p^{s-2}}-\frac{2}{p^{s-1}}-\frac{2}{p^s}} \\ &\le \frac{(1-\frac{1}{p})(s+1)(s+2)(p+1)^2}{2s(p^2-1)} = \frac{p+1}{2p}\cdot\frac{(s+1)(s+2)}{s}. \end{aligned}$$

The special values for $g(p^s)$ are the results of straightforward computations. □

Corollary 4.22. *Suppose that n is a positive integer with prime factorisation $n = p_1^{s_1}\cdot\dots\cdot p_t^{s_t}$. Then we have*

$$\frac{\mathcal{E}_{\max}(n)}{\tilde{\mathcal{E}}_{\max}(n)} \le \tau(n)\prod_{i=1}^{t}\frac{p_i+1}{2p_i}\prod_{\substack{i=1\\ 2\nmid s_i}}^{t}\left(1+\frac{1}{s_i+1}\right)\prod_{\substack{i=1\\ 2\mid s_i}}^{t}\left(1+\frac{2}{s_i}\right).$$

Proof. By Theorem 4.19 (i), Theorem 4.15, and Proposition 4.21,

$$\begin{aligned} \frac{\mathcal{E}_{\max}(n)}{\tilde{\mathcal{E}}_{\max}(n)} &\le \frac{nf(n)}{\tilde{\mathcal{E}}_{\max}(n)} = \prod_{i=1}^{t} g(p_i^{s_i}) \\ &\le \prod_{i=1}^{t}\frac{p_i+1}{2p_i}\prod_{\substack{i=1\\ 2\nmid s_i}}^{t}(s_i+2)\prod_{\substack{i=1\\ 2\mid s_i}}^{t}\frac{(s_i+1)(s_i+2)}{s_i}. \end{aligned} \tag{4.50}$$

Since $\tau(n) = \prod_{i=1}^{t}(s_i+1)$, the corollary follows. □

Proof of Theorem 4.19 (iii). We cannot use Corollary 4.22 directly, but by (4.50) we know that

$$\frac{\mathcal{E}_{\max}(n)}{\tilde{\mathcal{E}}_{\max}(n)} \le \prod_{p\in\mathbb{P},\, p\mid n} g(p^{e_p(n)}).$$

Since $\tau(n) = \prod_{p\in\mathbb{P},\, p\mid n}(e_p(n)+1)$, it suffices to show that $g(p^s) \le s+1$ for any prime power p^s. This will be verified by use of the different bounds obtained in Proposition 4.21.

Case 1: $2 \nmid s$.

We have $g(p^s) \le \frac{p+1}{2p}(s+2) \le s+1$ for all p and s except for $p = 2$, $s = 1$, but $g(2) = 2$ closes the gap.

Case 2: $s = 2$.

This case is settled by the fact that $g(p^2) \le 3$.

Case 3: $2 \mid s$ and $s \geq 4$.

Here we have $g(p^s) \leq \frac{p+1}{2p} \cdot \frac{(s+1)(s+2)}{s} \leq s+1$ for all p and s except for $p=2$, $s=4$, but we know that $g(2^4) = \frac{64}{17} \leq 5$. □

Besides multiplicativity, our proofs concerning $\mathcal{E}_{\min}(n)$ will be based on knowledge about the second largest modulus of the eigenvalues of $\mathrm{ICG}(n, \mathcal{D})$ (cf. Remark 4.17 (iii)), i.e., about

$$\Lambda(n, \mathcal{D}) := \max\{|\lambda| : \ \lambda \in \mathrm{Spec}(\mathrm{ICG}(n, \mathcal{D})), \ |\lambda| < \Phi(n, \mathcal{D})\}, \tag{4.51}$$

which we only define if $\mathrm{ICG}(n, \mathcal{D})$ has eigenvalues differing in modulus. It will be crucial for us to gather some facts about $\Lambda(n, \mathcal{D})$. The first step is

Lemma 4.23. *Let n be a positive integer and $\mathcal{D} \subseteq D(n)$ with $n \in \mathcal{D}$.*

(i) *Then $\mathcal{E}(n, D) \geq n$.*

(ii) $\mathrm{ICG}(n, \mathcal{D})$ *has a negative eigenvalue if and only if $\mathcal{E}(n, \mathcal{D}) > n$.*

Proof. From linear algebra we know that $\sum_{k=1}^{n} \lambda_k(n, \mathcal{D})$ equals the trace of the adjacency matrix of $\mathrm{ICG}(n, \mathcal{D})$. Since $n \in \mathcal{D}$ by hypothesis, $\mathrm{ICG}(n, \mathcal{D})$ has a loop at every vertex, i.e., all diagonal entries of its adjacency matrix are 1. Hence, $\sum_{k=1}^{n} \lambda_k(n, \mathcal{D}) = n$, and consequently

$$\mathcal{E}(n, \mathcal{D}) = \sum_{k=1}^{n} |\lambda_k(n, \mathcal{D})| \geq \left| \sum_{k=1}^{n} \lambda_k(n, \mathcal{D}) \right| = n, \tag{4.52}$$

which proves (i). Equality in (4.52) only holds if all $\lambda_k(n, \mathcal{D})$ have the same sign. We know that $\lambda_n(n, \mathcal{D}) = \Phi(n, \mathcal{D}) > 0$, and this implies (ii). □

Proposition 4.24. *Let p^s be a prime power and $\mathcal{D} \subseteq D(p^s)$ with $p^s \in \mathcal{D}$, and set $r := |\mathcal{D}|$.*

(i) *For $r = 1$, i.e., $\mathcal{D} = \{p^s\}$, we have* $\mathrm{Spec}(\mathrm{ICG}(n, \mathcal{D})) = \{1\}$.

(ii) *Let $r \geq 2$. Then $\mathcal{D}$ is uni-regular if and only if $\Lambda(p^s, \mathcal{D}) = 0$. In this case the maximal eigenvalue $\Phi(p^s, \mathcal{D}) = p^{s-a_1} = p^{r-1}$ has multiplicity p^{a_1}.*

(iii) *If $r \geq 2$ and $\mathcal{D}$ is not uni-regular, then $\mathcal{E}(p^s, \mathcal{D}) > p^s$.*

Proof. Let $\mathcal{D} = \{p^{a_1}, p^{a_2}, \ldots, p^{a_{r-1}}, p^{a_r}\}$ with $0 \leq a_1 < a_2 < \cdots < a_{r-1} < a_r = s$. Hence in case $r = 1$, that is $\mathcal{D} = \{p^s\}$, we trivially have $\lambda_k(p^s, \mathcal{D}) = c(k, 1) = 1$ for all k, which proves (i).

Henceforth we assume $r \geq 2$. On setting $j := e_p(k)$, we apply Lemma 4.4 and distinguish two cases.

Case 1: $s - a_\ell \leq j \leq s - a_{\ell-1} - 2$ for some $1 \leq \ell \leq r$, where $a_0 := -2$.

Using the notation $\mathcal{D}(x) := \{d \in \mathcal{D} : d \geq x\}$, we obtain from Lemma 4.4 that

$$\lambda_k(p^s, \mathcal{D}) = \sum_{\substack{i=1 \\ a_i \geq a_\ell}}^{r} \varphi(p^{s-a_i}) = \Phi(p^s, \mathcal{D}(p^{a_\ell})), \tag{4.53}$$

with $\Phi(n, \mathcal{D})$ as defined in (4.27).

Case 2: $j = s - a_\ell - 1$ for some $1 \leq \ell \leq r - 1$.

Now Lemma 4.4 yields

$$\lambda_k(p^s, \mathcal{D}) = \sum_{\substack{i=1 \\ a_i \geq a_\ell + 1}}^{r} \varphi(p^{s-a_i}) - p^{s-a_\ell-1} = \Phi(p^s, \mathcal{D}(p^{a_\ell+1})) - p^{s-a_\ell-1}. \tag{4.54}$$

In order to prove (ii), we first assume that $\mathcal{D} = \{p^{s-r+1}, p^{s-r+2}, \ldots, p^{s-1}, p^s\}$ is uni-regular. We observe that for $a_\ell = a_{\ell-1} + 1$ $(2 \leq \ell \leq s)$ the corresponding interval considered in Case 1 is empty. Hence Case 1 occurs only if $\ell = 1$, i.e., for $s - a_1 \leq j \leq s$, and then

$$\lambda_k(p^s, \mathcal{D}) = \Phi(p^s, \mathcal{D}(p^{a_1})) = \Phi(p^s, \mathcal{D}) = \lambda_{p^s}(p^s, \mathcal{D}), \tag{4.55}$$

which is the largest eigenvalue. This reflects the phenomenon that the largest eigenvalue has multiplicity greater than 1 if $a_1 > 0$, that is, the elements of $\mathcal{D}$ are not co prime or, equivalently, $\mathrm{ICG}(n, \mathcal{D})$ is disconnected (see Remark 4.17 (i)). More precisely, we have for each $j = s - u$, $0 \leq u \leq a_1$, exactly $\varphi(p^u)$ integers $k = p^j m$ with $p \nmid m$ and $1 \leq k \leq p^s$. Hence the multiplicity of the largest eigenvalue $\Phi(p^s, \mathcal{D})$ is precisely $\sum_{u=0}^{a_1} \varphi(p^u) = p^{a_1}$. By (4.55) we know that $\Phi(p^s, \mathcal{D}) = \Phi(p^s, \mathcal{D}(p^{a_1}))$, and since $a_1 = s - r + 1$ in $\mathcal{D} = \{p^{s-r+1}, p^{s-r+2}, \ldots, p^{s-1}, p^s\}$, the asserted formulas for $\Phi(p^s, \mathcal{D})$ in (ii) follow.

By the argument above, eigenvalues other than the largest one can only appear in Case 2. For $\mathcal{D} = \{p^{s-r+1}, p^{s-r+2}, \ldots, p^{s-1}, p^s\}$ and $j = r - \ell - 1$ $(1 \leq \ell \leq r)$, we obtain by (4.54)

$$\lambda_k(p^s, \mathcal{D}) = \Phi(p^s, \mathcal{D}(p^{s-r+\ell+1})) - p^{r-\ell-1} = \sum_{i=0}^{r-\ell-1} \varphi(p^i) - p^{r-\ell-1} = 0.$$

This proves $\Lambda(p^s, \mathcal{D}) = 0$.

To complete the proof of (ii), it remains to show that $\Lambda(p^s, \mathcal{D}) \neq 0$ for any non-regular set $\mathcal{D}$. Such a divisor set can be written as $\mathcal{D} = \{p^{a_1}, \ldots, p^{a_\ell}, p^{a_{\ell+1}}, \ldots, p^{a_r}\}$

with $0 \leq a_1 < a_2 < \cdots < a_r = s$, $a_{\ell+1} - a_\ell \geq 2$, and $a_{i+1} - a_i = 1$ for some $1 \leq \ell \leq r-1$ and all $i = \ell+1, \ldots, r-1$. Then for all $k = p^{s-a_\ell-1}m$, $p \nmid m$, i.e. $j = s - a_\ell - 1$, have by (4.54) in Case 2

$$\lambda_k(p^s, \mathcal{D}) = \Phi(p^s, \mathcal{D}(p^{a_\ell+1})) - p^{s-a_\ell-1} = \sum_{i=\ell+1}^{r} \varphi(p^{s-a_i}) - p^{s-a_\ell-1} < 0, \tag{4.56}$$

hence $\Lambda(p^s, \mathcal{D}) \neq 0$.

It remains to verify (iii). Since $\mathcal{D}$ is not uni-regular, we have negative eigenvalues by (4.56), and Lemma 4.23 (ii) proves $\mathcal{E}(p^s, \mathcal{D}) > p^s$. □

It was shown in [40], Theorem 3.1, that for a prime power p^s

$$\mathcal{E}_{\min}(p^s) = 2p^s\Big(1 - \frac{1}{p}\Big). \tag{4.57}$$

Observe that the minimum is extended over divisor sets $\mathcal{D} \subseteq D^*(p^s)$, i.e., over loopless graphs. Moreover, the p^s-minimal divisor sets were identified precisely as the singleton sets $\mathcal{D} = \{p^j\}$ with $0 \leq j \leq s-1$. For our purpose we shall require a corresponding result for graphs $\mathrm{ICG}(p^s, \mathcal{D})$ containing loops, that is with $p^s \in \mathcal{D}$.

Proposition 4.25. *Let p^s be a prime power. Then*

$$\hat{\mathcal{E}}_{\min}(p^s) := \min\{\mathcal{E}(p^s, \mathcal{D}) : \ p^s \in \mathcal{D} \subseteq D(p^s)\} = p^s, \tag{4.58}$$

where the minimising divisor sets are exactly the uni-regular ones.

The reader might notice that (4.57) implies $\hat{\mathcal{E}}_{\min}(p^s) \leq \mathcal{E}_{\min}(p^s)$, where equality only holds in case $p = 2$.

Proof of Proposition 4.25. For $r = 1$, the assertion follows immediately from Proposition 4.24(i). Hence assume that $r \geq 2$. By Proposition 4.24(iii), it suffices to show that we have $\mathcal{E}(p^s, \mathcal{D}) = p^s$ for each uni-regular divisor set $\mathcal{D} = \{p^{a_1}, p^{a_1+1}, \ldots, p^{s-1}, p^s\}$. We know from Proposition 4.24(ii) that $\mathrm{ICG}(n, \mathcal{D})$ has only two different eigenvalues, namely $\Phi(p^s, \mathcal{D}) = p^{s-a_1}$ with multiplicity p^{a_1} and 0 (consequently with multiplicity $p^s - p^{a_1}$). Therefore, $\mathcal{E}(p^s, \mathcal{D}) = p^{a_1} \cdot p^{s-a_1} = p^s$, as required. □

Proof of Theorem 4.16. Let $\mathcal{D} \subseteq D^*(n)$ be a multiplicative set such that $\mathcal{E}(n, \mathcal{D}) = \tilde{\mathcal{E}}_{\min}(n)$. Then $\mathcal{D} = \prod_{i=1}^{t} \mathcal{D}^{(i)}$ for certain divisor sets $\mathcal{D}^{(i)} \subseteq D(p_i^{s_i})$ $(1 \leq i \leq t)$, and $\tilde{\mathcal{E}}_{\min}(n) = \prod_{i=1}^{t} \mathcal{E}(p_i^{s_i}, \mathcal{D}^{(i)})$ by [28], Corollary 4.1(ii). By the minimality of $\mathcal{E}(n, \mathcal{D})$ it follows from [40], Theorem 3.1, and our Proposition 4.25 that for $1 \leq i \leq t$

$$\mathcal{E}(p_i^{s_i}, \mathcal{D}^{(i)}) = \begin{cases} 2p_i^{s_i}(1 - \frac{1}{p_i}), & \text{if } p_i^{s_i} \notin \mathcal{D}^{(i)}, \\ p_i^{s_i}, & \text{if } p_i^{s_i} \in \mathcal{D}^{(i)}, \end{cases}$$

where $\mathcal{D}^{(i)}$ is either a singleton set $\{p_i^{u_i}\}$ for some $0 \leq u_i \leq s_i - 1$, or a uni-regular set containing $p_i^{s_i}$, respectively. This yields

$$\tilde{\mathcal{E}}_{\min}(n) = n \prod_{\substack{i=1 \\ p_i^{s_i} \notin \mathcal{D}^{(i)}}}^{t} 2\Big(1 - \frac{1}{p_i}\Big), \tag{4.59}$$

and our assumption $n \notin \mathcal{D}$ implies that $p_i^{s_i} \notin \mathcal{D}^{(i)}$ for at least one i. Under this restriction, and since p_1 is the smallest of the primes involved, it is easily seen that the right-hand side of (4.59) becomes minimal if $p_1^{s_1} \notin \mathcal{D}^{(1)}$ and $p_i^{s_i} \in \mathcal{D}^{(i)}$ for $2 \leq i \leq t$ with the corresponding divisor sets $\mathcal{D}^{(i)}$ just mentioned. □

We confined our study of the energies of integral circulant graphs to the rather restricted class having multiplicative divisor sets. Yet, somewhat unexpectedly, this led to good bounds for $\mathcal{E}_{\min}(n)$ and $\mathcal{E}_{\max}(n)$. On top of that, the results obtained by the study of multiplicative divisor sets, combined with some numerical evidence, encourage us to make the following two conjectures.

Conjecture 4.26. *For each integer $n \geq 2$, we have $\mathcal{E}_{\min}(n) = 2n\big(1 - \frac{1}{p_1}\big)$, where p_1 denotes the smallest prime factor of n.*

Conjecture 4.27. *Let $n \geq 2$ be an arbitrary integer. Then $\mathcal{E}(n, \mathcal{D}) = \mathcal{E}_{\min}(n)$ implies that $\mathcal{D}$ is a multiplicative divisor set.*

Observe that Conjecture 4.26 is a consequence of Conjecture 4.27 by Theorem 4.16.

In 2005 it was conjectured by So that, given a positive integer n, two graphs $\mathrm{ICG}(n, \mathcal{D}_1)$ and $\mathrm{ICG}(n, \mathcal{D}_2)$ are *cospectral*, that is $\mathrm{Spec}(\mathrm{ICG}(n, \mathcal{D}_1)) = \mathrm{Spec}(\mathrm{ICG}(n, \mathcal{D}_2))$, if and only if $\mathcal{D}_1 = \mathcal{D}_2$. For $n = p^s$ this follows immediately from (4.27) by straightforward comparison of the largest eigenvalues $\Phi(p^s, \mathcal{D}_1)$ and $\Phi(p^s, \mathcal{D}_2)$ of the two graphs. So's conjecture was also confirmed for the slightly more general case $n = p^s q^t$ with primes $p \neq q$ and $t \in \{0, 1\}$ (cf. [10] for details), but its proof required the study of eigenvalues other than the largest one as well as their multiplicities. For arbitrary positive integers n and arbitrary divisor sets the problem is still open. Therefore, we suggest to study the following weaker conjecture, which might be more accessible.

Conjecture 4.28. *Let n be a positive integer, and let $\mathcal{D}_1, \mathcal{D}_2 \subseteq D^*(n)$ be two multiplicative sets. If* $\mathrm{ICG}(n, \mathcal{D}_1)$ *and* $\mathrm{ICG}(n, \mathcal{D}_2)$ *are cospectral, then $\mathcal{D}_1 = \mathcal{D}_2$.*

Bibliography

[1] O. Ahmadi, N. Alon, I. F. Blake and I. E. Shparlinski, Graphs with integral spectrum. *Linear Algebra Appl.* 430 (2009), 547–552.

[2] N. Alon, Eigenvalues and expanders. *Combinatorica*, 6 (1986), 83–96.

[3] R. J. Angeles-Canul, R. Norton, M. Oppermann, C. Paribello, M. Russel and C. Tamon, Perfect state transfer, integral circulants and join of graphs. *Quantum Inf. Comput.* 10 (2010), 325–342.

[4] T. M. Apostol, *Introduction to Analytic Number Theory*, Springer Verlag, Berlin – Heidelberg – New York, 1976.

[5] H. Bass, The Ihara-Selberg zeta function of a tree lattice. *Intl. J. Math.* 3 (1992), 717–797.

[6] F. Bien, Construction of telephone networks by group representations. *Notices Am. Math. Soc.* 36 (1989), No. 1, 5–22.

[7] N. L. Biggs, *Algebraic Graph Theory*. 2nd edition, Cambridge University Press, Cambridge, 1993.

[8] R. B. Boppana, Eigenvalues and graph bisection: an average case analysis. In *Proc. 28th Ann. Symp. Found. Comp. Sci.*, IEEE 1987, 280–285.

[9] R. A. Brualdi, *Energy of a graph*. AIM Workshop Notes, 2006.

[10] C. F. Cusanza, *Integral circulant graphs with the spectral Ádám property*. Master's Thesis. Paper 2875. http://scholarworks.sjsu.edu/etd_theses/2875

[11] P. J. Davis, *Circulant Matrices*. John Wiley & Sons, New York – Chichester – Brisbane, 1979.

[12] R. Diestel, *Graph Theory*. Springer-Verlag, New York, 1997.

[13] A. Droll, A classification of Ramanujan unitary Cayley graphs. *Electron. J. Combin* 17 (2010), #N29.

[14] C. Godsil and G. Royle, *Algebraic Graph Theory*. Vol. 207 of Graduate Texts in Mathematics, Springer-Verlag, New York, 2001.

[15] I. Gutman, *The energy of a graph*. Ber. Math.-Stat. Sekt. Forschungszent. Graz 103, 1978.

[16] F. Harary and A. J. Schwenk, Which graphs have integral spectra? In *Graphs and Combinatorics* (Ed. R. Bari and F. Harary). Berlin: Springer-Verlag, pp. 45–51, 1974.

[17] G. H. Hardy, E. M. Wright and J. H. Silverman, *An Introduction to the Theory of Numbers*. Oxford University Press, Oxford, 2008.

[18] K. Hashimoto, Zetas functions of finite graphs and representations of p-adic groups. *Advanced Studies in Pure Math.*, vol. 15, Academic Press, New York, 1989, 211–280.

[19] Sh. Hoory, N. Linial and A. Wigderson, Expander graphs and their applications. *Bull. Amer. Math. Soc.* 43 (2006), no. 4, 439–561.

[20] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[21] Y. Ihara, On discrete subgroups of the two by two projective linear group over p-adic fields. *J. Math. Soc. Japon.* 18 (1966), 219–235.

[22] A. Ilić, Distance spectra and distance energy of integral circulant graphs. *Linear Algebra Appl.* 433 (2010), 1005–1014.

[23] A. Ivic, *The Riemann Zeta Function: The Theory of the Riemann Zeta-Function with Applications*. John Wiley & Sons Inc., New York, 1985.

[24] H. Iwaniec and E. Kowalski, *Analytic Number Theory*. AMS Publications, Providence, 2004.

[25] M. Kotani and T. Sunada, Zeta functions of finite graphs. *J. Math. Sci. Univ. Tokyo* 7 (2000), 7–25.

[26] T. Y. Lam and K. H. Leung On vanishing sums of roots of unity. *J. Algebra* 224 (2000), 91–109.

[27] S. Lang, *Algebraic Number Theory*. Addison-Wesley, Reading MA, 1968.

[28] T. A. Le and J. W. Sander, Convolutions of Ramanujan sums and integral circulant graphs. *Intl. J. Number Theory* 7 (2012), 1777–1788.

[29] T. A. Le and J. W. Sander, Extremal energies of integral circulant graphs via multiplicativity. *Lin. Algebra Appl.* 437 (2012), 1408–1421.

[30] T. A. Le and J. W. Sander, Integral circulant Ramanujan graphs of prime power order. *Electronic J. Combinatorics* 20(3) (2013), #P35, 12pp.

[31] A. Lubotzky, R. Phillips and P. Sarnak, Ramanujan graphs. *Combinatorica* 8 (1988), 261–277.

[32] R. B. Mallion, Some chemical applications of the eigenvalues and eigenvectors of certain finite, planar graphs. In *Applications of Combinatorics*, R. J. Wilson (Ed.), Shiva Publishing Ltd., Nantwich, Cheshire (England, United Kingdom), 1982, 87–114.

[33] S. J. Miller and R. Takloo-Bigash, *An Invitation to Modern Number Theory*. Princeton University Press, Princeton, 2006.

[34] M. Morgenstern, Existence and explicit constructions of $q+1$ regular Ramanujan graphs for every prime power q. *J. Combinatorial Theory*, Series B 62 (1994), 44–62.

[35] M. R. Murty, Ramanujan graphs. *J. Ramanujan Math. Soc.* 18 (2003), 33–52.

[36] W. Narkiewicz, On a class of arithmetical convolutions. *Colloq. Math.* 10 (1963), 81–94.

[37] S. J. Patterson, *An Introduction to the Theory of the Riemann Zeta Function*. Cambridge University Press, Cambridge, 1988.

[38] M. Rosen, *Number Theory in Function Fields*. Springer-Verlag, New York, 2002.

[39] J. W. Sander, Integral circulant Ramanujan graphs via multiplicativity and ultrafriable integers, *Linear Algebra Appl.* 477 (2015), 21–41.

[40] J. W. Sander and T. Sander, The energy of integral circulant graphs with prime power order. *App. Anal. Discrete Math.* 5 (2011), 22–36.

[41] J. W. Sander and T. Sander, The exact maximal energy of integral circulant graphs with prime power order. *Contr. Discr. Math.* 8 (2013), 19–40.

[42] N. Saxena, S. Severini and I. E. Shparlinski, Parameters of integral circulant graphs and periodic quantum dynamics. *Int. J. Quantum Inf.* 5 (2007), 417–430.

[43] J.-P. Serre, *Trees*. Springer-Verlag, New York, 1980.

[44] W. Schwarz and J. Spilker, *Arithmetical Functions*. Lond. Math. Soc., Lect. Notes Math. 184, Cambridge University Press, 1994.

[45] I. Shparlinski, On the energy of some circulant graphs. *Linear Algebra Appl.* 414 (2006), 378–382.

[46] W. So, Integral circulant graphs. *Discrete Math.* 306 (2005), 153–158.

[47] T. Sunada, Riemannian coverings and isospectral manifolds. *Ann. Math.* 121 (1985), 169–186.

[48] A. Terras, *Zeta Functions of Graphs*. Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 2011.

Index

Kathrin Bringmann
Yann Bugeaud
Titus Hilberdink
Jürgen Sander

Four Faces of Number Theory

This book arises from courses given at the International Summer School organized in August 2012 by the number theory group of the Department of Mathematics at the University of Würzburg. It consists of four essentially self-contained chapters and presents recent research results highlighting the strong interplay between number theory and other fields of mathematics, such as combinatorics, functional analysis and graph theory. The book is addressed to (under)graduate students who wish to discover various aspects of number theory. Remarkably, it demonstrates how easily one can approach frontiers of current research in number theory by elementary and basic analytic methods.

Kathrin Bringmann gives an introduction to the theory of modular forms and, in particular, so-called Mock theta-functions, a topic which had been untouched for decades but has obtained much attention in the last years. Yann Bugeaud is concerned with expansions of algebraic numbers. Here combinatorics on words and transcendence theory are combined to derive new information on the sequence of decimals of algebraic numbers and on their continued fraction expansions. Titus Hilberdink reports on a recent and rather unexpected approach to extreme values of the Riemann zeta-function by use of (multiplicative) Toeplitz matrices and functional analysis. Finally, Jürgen Sander gives an introduction to algebraic graph theory and the impact of number theoretical methods on fundamental questions about the spectra of graphs and the analogue of the Riemann hypothesis.

ISBN 978-3-03719-142-2

9 783037 191422

www.ems-ph.org